1977

Non-homogeneous Markov chains and their applications

Cheng-Chi Huang
Iowa State University

77-29,843

HUANG, Cheng-Chi, 1941-
NON-HOMOGENEOUS MARKOV CHAINS
AND THEIR APPLICATIONS.

Iowa State University, Ph.D., 1977
Statistics

Xerox University Microfilms, Ann Arbor, Michigan 48106

Non-homogeneous Markov chains and their applications

by

Cheng-Chi Huang

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of

The Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Co-majors: Mathematics
Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Departments

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa
1977

TABLE OF CONTENTS

I. DEFINITIONS, REVIEW AND INTRODUCTION

A. Definitions and Notation

Definition I.A.1: A stochastic process $\{X_k\}$, $k = 1,2,\cdots$ with state space $S = \{1,2,3,\cdots\}$ is said to satisfy the Markov property if for every n and all states $i_1, i_2, \cdots, i_n$ it follows that

$$P[X_n = i_n \mid X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \cdots, X_1 = i_1]$$

$$= P[X_n = i_n \mid X_{n-1} = i_{n-1}].$$

Notation: $P[X_n = i_n \mid X_{n-1} = i_{n-1}] = p^{(n-1,n)}_{i_{n-1}, i_n}$.

Definition I.A.2: A discrete time Markov chain is said to be stationary or homogeneous in time if the probability of going from one state to another is independent of the time at which the step is being made. That is, for all states i, j,

$$P[X_n = j \mid X_{n-1} = i] = P[X_{n+k} = j \mid X_{n+k-1} = i]$$

for $k = -(n-1), -(n-2), \cdots, -1, 0, 1, 2, \cdots$. The Markov chain is said to be non-stationary or non-homogeneous if the condition for stationarity fails. For stationary chains the following notation is used:

$$P[X_{n+k} + j \mid X_{n+k-1} + i] = p_{ij}.$$

Note: (The terms non-stationary or non-homogeneous will be used interchangeably in this thesis.)

Definition I.A.3: The stochastic matrix of one step-transition probabilities from time t to $t+1$, which we denote by P_{t+1}, is defined for $t \geq 0$ to be:

$$P_{t+1} = \begin{bmatrix} p_{1i}^{(t,t+1)} & p_{12}^{(t,t+1)} & \cdots \\ p_{21}^{(t,t+1)} & p_{22}^{(t,t+1)} & \cdots \\ \vdots & \vdots & \\ p_{m1}^{(t,t+1)} & p_{m2}^{(t,t+1)} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

where for $t \geq 0$:

$$p_{ij}^{(t,t+1)} \geq 0, \quad \forall i,j \in S$$

and

$$\sum_{j \in S} p_{ij}^{(t,t+1)} = 1, \quad \forall i \in S.$$

Definition I.A.4: For a stationary Markov chain, a subset C, of the state space S, is called closed if $p_{ik} = 0$ for all $i \in C$ and $k \notin C$.

Definition I.A.5: A stationary Markov chain is called irreducible if there exists no closed set other than S itself. If S has a proper closed subset, the chain is called reducible.

Definition I.A.6: Two states i and j are said to intercommunicate if for some $n \geq 0$, $p_{ij}^{(n)} > 0$ and for some $m \geq 0$, $p_{ji}^{(m)} > 0$.

It is known that a Markov chain is irreducible if and only if any two states intercommunicate.

Example I.A.1: Let $\{X_n\}$ be a discrete time stationary Markov chain with state space $\{1,2,3\}$ and transition matrix

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

then $p_{12}^{(1)} > 0$, $p_{21}^{(2)} > 0$, $p_{13}^{(2)} > 0$, $p_{31}^{(1)} > 0$, $p_{23}^{(1)} > 0$, $p_{32}^{(2)} > 0$, (note that: $p_{ij}^{(n)}$ = probability of going from state i to state j in n steps.) Therefore by definition we know that the chain is irreducible.

Example I.A.2: Let $\{X_n\}$ be a discrete time stationary Markov chain with state space $\{1,2,3\}$ and transition matrix

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \text{ then } C_1 = \{1,2\},\ C_2 = \{3\}$$

are closed subsets of S.

Definition I.A.7: For a stationary Markov chain, state j has period d if the following two conditions hold.

i) $p_{jj}^{(n)} = 0$ unless $n = md$ for some positive integer m and

ii) d is the largest integer with property i). State j is called aperiodic when $d = 1$.

(Recall that $p_{jj}^{(n)}$ = probability of going from state j to state j in n steps.)

Example I.A.3: Let $\{X_n\}$ be a discrete time stationary Markov chain with state space $\{1,2,3,4\}$ and transition matrix

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

For this chain $p_{11}^{(n)} = 0$ unless n is even. So the period of state 1 is even. Condition i) is satisfied by the even integers 2 and 4. Hence by condition ii) of the definition, the period of the state 1 is 4.

It is known that state j has period λ if and only if λ is the greatest common division of all those n's for which $p_{jj}^{(n)} > 0$. (i.e. $\lambda = \text{G.C.D.}\{n : p_{jj}^{(n)} > 0\}$.)

Example I.A.4: Let $\{X_n\}$ be a stationary Markov chain with state space $\{1,2,3\}$ and transition matrix

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} .$$

For this chain $p_{11}^{(1)} = \frac{1}{2} > 0$, therefore state 1 is aperiodic.

Definition I.A.8: For a stationary Markov chain, let $f_{ij}^{(n)}$ denote the probability that the first visit to state j from state i occurs at time n, that is,

$$f_{ij}^{(n)} = P[X_{n+k} = j, X_{n+k-1} \neq j, X_{n+k-2} \neq j, \cdots, X_{k+1} \neq j \mid X_k = i] .$$

If $i = j$ we refer to $f_{ii}^{(n)}$ as the probability that the first return to state i occurs at time n. By definition we say $f_{ij}^{(0)} = f_{ii}^{(0)} = 0$.

<u>Example I.A.5</u>: Let $\{X_n\}$ be a Markov chain with state space, $\{1,2\}$ and transition matrix

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix} .$$

For this example $f_{11}^{(2)} = \frac{1}{2}$ and $f_{22}^{(2)} = \frac{1}{2}$. Note that $p_{11}^{(2)} = \frac{3}{4}$ and $p_{22}^{(2)} = \frac{1}{2}$ so that p's and f's need not be the same.

<u>Definition I.A.9</u>: For a stationary Markov chain, we define

$$f_{ij}^{*} = \sum_{n=1}^{\infty} f_{ij}^{(n)} = \text{the probability of ever visiting state } j$$

from state i. A state, j, is said to be persistent if $f_{jj}^{*} = 1$. Otherwise it is called a transient state. Furthermore

i) if $\mu_j = \sum_{n=1}^{\infty} nf_{jj}^{(n)} = \infty$ and $f_{jj}^* = 1$ then the state j is called null persistent.

ii) if $\mu_j = \sum_{n=1}^{\infty} nf_{jj}^{(n)} < \infty$ and $f_{jj}^* = 1$ then the state j is called positive persistent.

<u>Example I.A.6</u>: Consider the Markov chain of Example I.A.5. States 1 and 2 are persistent, since

$$f_{11}^* = f_{11}^{(1)} + f_{11}^{(2)} = \frac{1}{2} + \frac{1}{2} = 1,$$

$$f_{22}^* = f_{22}^{(1)} + f_{22}^{(2)} + \cdots + f_{22}^{(k)} + \cdots = 0 + \frac{1}{2} + \frac{1}{4} + \cdots = 1,$$

and

$$\mu_1 = \sum_{k=1}^{\infty} kf_{11}^{(k)} = (1)\left(\frac{1}{2}\right) + (2)\left(\frac{1}{2}\right) = \frac{3}{2},$$

$$\mu_2 = \sum_{k=2}^{\infty} kf_{22}^{(k)} = \sum_{k=2}^{\infty} \frac{k}{2^{k-1}} = 3.$$

Hence both of states 1 and 2 are positive persistent. It is known that if the state space is finite, then all the persistent states are positive persistent.

Definition I.A.10: Let $P = (p_{ij})$ be the transition matrix for a stationary Markov chain. If $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$ exists for all j independently of i and if $\sum_{j=1}^{\infty} \pi_j = 1$, then we say the chain is ergodic.

Example I.A.7: Let $\{X_n\}$ be a Markov chain with state space $S = \{1,2,3,4\}$ and transition matrix,

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} \end{pmatrix}.$$

It is easy to show that this chain is aperiodic but not irreducible. In this case $p_{11}^{(n)} \equiv 1$ for all n, $p_{21}^{(n)} \equiv 0$ for all n, $p_{31}^{(n)} = \frac{1}{3}$ for all n, and

$\lim_{n\to\infty} p_{41}^{(n)} = \frac{1}{2}$. Hence the limit of $p_{i1}^{(n)}$ exists but it certainly depends on state i.

<u>Example I.A.8</u>: Let $\{X_n\}$ be a Markov chain with state space, $S = \{1,2\}$ and transition matrix, $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. It is easy to see that this chain is irreducible but not aperiodic. In this case $p_{11}^{(n)} = 0$ if n is odd and 1 if n is even. Hence $\lim_{n\to\infty} p_{11}^{(n)}$ does not exist.

In view of the above two examples it can be shown that the following conditions are necessary for ergodicity.

i) All the persistent states are aperiodic.

ii) There is at most one irreducible closed subset of persistent states.

<u>Example I.A.9</u>: Let $\{X_n\}$ be a Markov chain with state space $S = \{1,2,3,4\}$ and transition matrix,

$$P = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix} .$$

Then $p_{i1}^{(n)} \to \frac{4}{13}$, $p_{i2}^{(n)} \to \frac{9}{13}$, $p_{i3}^{(n)} \to 0$, $p_{i4}^{(n)} \to 0$ independently of i $(i = 1,2,3,4)$. Thus the chain is ergodic.

It can be shown that a finite Markov chain is ergodic if and only if there is exactly one irreducible closed subset of positive persistent states and all these states are aperiodic. But if a Markov chain with one irreducible aperiodic positive persistent class has infinitely many states, it may not be ergodic

Example I.A.10: Let $\{X_n\}$ be a stationary Markov chain with state space $\{1,2,3\cdots\}$ and transition matrix,

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 0 & 0 & \cdots \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 1 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} .$$

In this example

$$\begin{aligned} f^*_{11} &= f_{11}^{(1)} + f_{11}^{(2)} + \cdots \\ &= 1 + 0 + 0 + 0 + \cdots \\ &= 1, \end{aligned}$$

and
$$\begin{aligned} \mu_1 &= (1)f_{11}^{(1)} + (2)f_{11}^{(2)} + \cdots \\ &= (1)(1) + (2)(0) + 0 + 0 + \cdots \\ &= 1 < \infty. \end{aligned}$$

Thus $C = \{1\}$ is the only irreducible closed subset of positive persistent states (for states $2,3,4\cdots$ are transient states) and state 1 is aperiodic. Now states $2,3,4,5\cdots$ are all transient, and the chain never enters the irreducible closed set $C = \{1\}$ from the transient states $\{4,5,6\cdots\}$. Therefore the chain is not ergodic.

<u>Definition I.A.12</u>: Let $P_1, P_2, \cdots$ be transition matrices for a non-stationary Markov chain with starting vector $f^{(0)}$, (i.e. $f^{(0)} = (f_1^{(0)}, f_2^{(0)}, f_3^{(0)}, \cdots)$, $f_i^{(0)} \geq 0$ $(i = 1,2,3,\cdots)$ and $\sum_{i=1}^{\infty} f_i^{(0)} = 1)$. Define

$$f^{(k)} = f^{(0)} P_1 \cdot P_2 \cdot \ldots \cdot P_k$$

and

$$f^{(m,k)} = f^{(0)} P_{m+1} \cdot P_{m+2} \cdot \ldots \cdot P_k .$$

Define the $j^{\underline{th}}$ element of $f^{(k)}$ by $f_j^{(k)}$ and define the $(i,j)^{\underline{th}}$ element of $P^{(m,k)} = P_{m+1} \cdot P_{m+2} \cdot \ldots \cdot P_k$ by

$p_{ij}^{(m,k)}$.

Definition I.A.13: If $f = (f_1, f_2, f_3 \cdots)$ is a vector, define the norm of f by

$$\|f\| = \sum_{i=1}^{\infty} |f_i|.$$

If $A = (a_{ij})$ is a matrix, define the norm of A by

$$\|A\| = \sup_i \sum_{j=1}^{\infty} |a_{ij}|.$$

(Note: There are many norms that could be placed upon spaces of vectors and matrices of countable dimension, but we will consider only the above norm on each of these spaces.) The norm is used to give the following definitions of weak and strong ergodicity.

Definition I.A.14: A non-stationary Markov chain $\{P_n\}$ is called weakly ergodic if for all m,

$$\lim_{k\to 0} \sup_{f^{(0)},g^{(0)}} \|f^{(m,k)} - g^{(m,k)}\| = 0.$$

where $f^{(0)}$ and $g^{(0)}$ are starting vectors.

Definition I.A.15: A non-stationary Markov chain $\{P_n\}$ is called strongly ergodic if there exists a vector $q = (q_1, q_2, \cdots)$, with $\|q\| = 1$ and $q_i \geq 0$ for all i, such that for all m,

$$\lim_{k\to\infty} \sup_{f^{(0)}} \|f^{(m,k)} - q\| = 0,$$

where $f^{(0)}$ is a starting vector.

From the above definitions we see that strong ergodicity implies weak ergodicity. In particular for a weakly ergodic chain we have that the effect of the initial distribution is lost so that $f^{(m,k)}$ and $g^{(m,k)}$ are close to each other in the sense of the "norm", but they are not necessarily close to any fixed vector. Such behavior will be referred to as "loss of memory". However, by Definition I.A.14 strong ergodicity implies convergence and loss of memory. It can be shown that a non-stationary Markov chain with transition matrices $\{P_n\}$ is strongly ergodic if and only if there exists a row-constant matrix Q such that for each m,

$$\lim_{k\to\infty} \| P^{(m,k)} - Q \| = 0.$$

Example I.A.11: Let $\{X_n\}$ be a non-stationary Markov chain with transition matrices

$$P_{2n-1} = \begin{pmatrix} 1 - \frac{1}{2n-1}, & \frac{1}{2n-1} \\ 1 - \frac{1}{2n-1}, & \frac{1}{2n-1} \end{pmatrix}, \quad P_{2n} = \begin{pmatrix} \frac{1}{2n}, & 1 - \frac{1}{2n} \\ \frac{1}{2n}, & 1 - \frac{1}{2n} \end{pmatrix}$$

for $n = 1, 2, \cdots$. Then, for any starting vector $f^{(0)}$, we have

$$f^{(m,k)} = \left(1 - \frac{1}{k}, \frac{1}{k}\right) \quad \text{if } k \text{ is odd}$$

$$= \left(\frac{1}{k}, 1 - \frac{1}{k}\right) \quad \text{if } k \text{ is even,}$$

so the chain is weakly ergodic but not strongly ergodic.

Example I.A.12: Let $\{X_n\}$ be a non-stationary Markov chain with transition matrices

$$P_n = \begin{pmatrix} \frac{1}{2} - \frac{1}{n+1}, & \frac{1}{2} + \frac{1}{n+1} \\ \frac{1}{2} - \frac{1}{n+1}, & \frac{1}{2} + \frac{1}{n+1} \end{pmatrix}$$

for $n = 1,2,3,\cdots$. Then, for any starting vector $f^{(0)}$, we have $f^{(m,k)} = \left(\frac{1}{2} - \frac{1}{k+1}, \frac{1}{2} + \frac{1}{k+1}\right) \to \left(\frac{1}{2}, \frac{1}{2}\right)$, as $k \to \infty$ so the chain is strongly ergodic.

In general, it is difficult to show that a chain is weakly or strongly ergodic by using the definition directly. In view of this difficulty we will introduce the ergodic coefficient as a measure of how close a stochastic matrix is to having identical rows. The ergodic coefficient as defined below was introduced by Dobrushin.

Definition I.A.16: Let P be a stochastic matrix. The ergodic coefficient of P, denoted by $\alpha(P)$, is defined by

$$\alpha(P) = 1 - \sup_{i,k} \sum_{j=1}^{\infty} [p_{ij} - p_{kj}]^+,$$

where $[p_{ij} - p_{kj}]^+ = \max(0, p_{ij} - p_{kj})$.

It is sometimes more convenient to use $1 - \alpha(P)$ instead of $\alpha(P)$ itself. In view of this we define

$$\delta(P) = 1 - \alpha(P),$$

and call $\delta(P)$ the delta coefficient of P.

It can be shown that

$$\sup_{i,k} \sum_{j=1}^{\infty} [p_{ij} - p_{kj}]^+ = \frac{1}{2} \sup_{i,k} \sum_{j=1}^{\infty} |p_{ij} - p_{kj}|.$$

Therefore

$$\alpha(P) = 1 - \frac{1}{2} \sup_{i,k} \sum_{j=1}^{\infty} |p_{ij} - p_{kj}|$$

and

$$\delta(P) = \frac{1}{2} \sup_{i,k} \sum_{j=1}^{\infty} |p_{ij} - p_{kj}|.$$

Example I.A.13: Let

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

then

$$\alpha(P) = 1 - \frac{1}{2} \sup_{i,k} \sum_{j=1}^{\infty} |p_{ij} - p_{kj}|$$

$$= 1 - \frac{1}{2} \max\left\{1 + \frac{1}{2} + \frac{1}{2}, 1 + 0 + 0, 1 + \frac{1}{2} + \frac{1}{2}\right\}$$

$$= 1 - \frac{1}{2}(2)$$

$$= 0,$$

and $\delta(P) = 1 - \alpha(P) = 1.$

Note that the ergodic coefficient can be used to determine whether or not a non-homogeneous Markov chain is weakly ergodic. In many cases the following theorem will be easier to use than the original definition.

Theorem: A non-stationary Markov chain is weakly ergodic if and only if for all m, $\delta(P^{(m,k)}) \to 0$ as $k \to \infty$.

Note: More definitions will be introduced in Chapter III.

B. Review and Introduction

This dissertation contains two main parts, the first (Chapter II and III) establishes new results in the theory of non-stationary Markov chains and the second (Chapter IV) considers some applications of non-stationary Markov chains.

In Chapter II we use the ergodic coefficient to prove that if P is strongly ergodic, then P^n converges to a row-constant matrix at a geometric rate (Lemma II.1); also in Lemma II.2 we use the ergodic coefficient to show that if P_n converges to P with P strongly ergodic, then the tail of the non-stationary Markov chain P_n has properties similar to those of P. The ergodic coefficient was defined by Dobrushin (1956), who showed why it was natural in the theory of Markov chains to make use of the ergodic coefficient. Ever since Dobrushin gave a sufficient condition for the (weak) ergodicity of a non-stationary discrete time Markov chain, the ergodic coefficient of a stochastic matrix has been used to great advantage in the study of non-stationary chains by many authors. In an article on ergodic properties of non-stationary finite Markov chains, Hajnal (1956) dealt with the behavior of finite non-stationary Markov chains having regular transition matrices. Two types of ergodic behavior (weak and strong ergodicity) were distinguished, and sufficient conditions given for each type. But Mott (1957) and Hajnal (1956), apparently unaware of Dobrushin's work, both implicitly required conditions in terms of the ergodic coefficient for a non-stationary finite Markov chain to be weakly or strongly ergodic. Since Mott's and Hajnal's works were limited to the non-stationary finite Markov chains, Paz (1970) extended the work of Hajnal to infinite state space using the ergodic coefficient to determine the strong or weak ergodicity of a non-stationary chain. One of the important consequences, used in the proof of the theorem in Chapter II, is that, according to the definitions of weak and strong ergodicity, it is true that a weakly ergodic chain need not be strongly ergodic, and it is possible to use the properties of weak ergodicity to show that a chain is strongly ergodic. Mott (1957) proved that if P_n is a non-stationary finite Markov chain which converges to P and P is weakly ergodic, then the chain is strongly ergodic. Later on, it was proved Bowerman, David and Isaacson (1975) that the result is true if the chain is finite or countably infinite. Besides Dobrushin's conditions in terms of ergodic coefficient and Mott's and Hajnal's implicit conditions in terms of the ergodic coefficient for a chain to be ergodic, Conn (1969), Madsen and Conn (1973), and Madsen and Isaacson (1973) gave conditions for ergodicity in terms of left eigenvectors. In Chapter III we use mean visit time as a criterion for a chain to be ergodic.

After much discussion of weak and strong ergodicity, Pullman and Styan (1973) worked on the convergence of Markov chains with non-stationary transition probabilities and constant causative matrix. A Markov chain P_n which satisfies $P_{n+1} = P_n C$ for all $n \geq 1$ was said to have constant causative matrix C (matrix C need not be stochastic). Lipstein (1965) suggested that if $\lim C^n = el'$ for some l', then n-step transition matrices $T_n = P_1P_2P_3\cdots P_n$ would also converges to el' as n approaches infinity. This was proved for 2-state chains only; and the methods used did not seem to apply to chains with more states. Pullman and Styan did prove that the T_n indeed converges to el' for chains with any number of states; in fact they converge so rapidly that $\sum_n \|T_n - el'\|$ converges. This is true even if the chain has countably many states, provided only $\sup_j \left\{ \sum_j |c_{ij}| \right\} < \infty$. More precisly, Pullman and Styan (1973) showed that if the powers of such a causative matrix C converge to L and $P_1L = L$ then $\sum_n \|T_n - L\|$ converges. We will see that Lipstein's (1965) conjecture is a simple consequence of Corollary II.8. This corollary says that if $P_n = P_1C^{n-1}$ where C is of infinite order, causative with respect to P_1 and bounded and if C^n converges to L where L is a row-constant stochastic matrix, then $P^{(t,t+n)}$ converges to L uniformly in t at a geometric rate.

Later Pitman (1974) worked on the uniform rate of convergence for Markov chain transition probabilities. Let P be a Markov matrix with a countable set of states J, and denote the n-th iterate of P by

$$P^n = \left(p_{ij}^{(n)}\right) \quad i,j \in J, \quad n = 0,1,2,\cdots.$$

If P is irreducible and aperiodic, then, according to a well-known theorem of Kolmogorov

$$\lim_{n\to\infty} p_{ij}^{(n)} = 1/m_{jj}$$

where m_{jj} denotes the expected recurrence time of state j and $1/m_{jj}$ is taken to be zero in the transient and null persistent cases when $m_{jj} = \infty$. Breiman gave an elegant proof of this theorem in the positive persistent case, based on the simple probabilites device of comparing the progress of two independent Markov chains with the same transition probabilities P but different initial distributions. Pitman's paper (1974) shows that this device can in fact be used to cover the null persistent case, too, and more importantly that in the positive persistent case it is possible to further exploit the idea to obtain a number of new and powerful refinements of the main limit theorem. The central results of the paper provided new results on the rate of convergence of $\lim_{n\to\infty} p_{ij}^{(n)} = 1/m_{jj}$ for Markov chain with infinite state space. These results generalize the work of Feller and Karlin on the rate of convergence of renewal sequences. Yet Pitman gives no information about the non-stationary case. Therefore in this dissertation (Theorem II.7) we determine the rate of convergence of the product $P_{m+1}P_{m+2}\cdots P_{m+n}$ to a row-constant matrix Q in terms of the rate of convergence of P_n to P where P is strongly ergodic. Also in the article on the uniform rates of convergence for Markov chain transition probabilities, Pitman (1974) used mean visit times to discuss the uniform rate of convergence of a Markov chain. Pitman pointed out that an infinite ergodic chain may have infinite mean visit time for certain starting vectors. In this dissertation we use the finiteness of the mean visit time to state j in the characterization of uniform strong ergodicity for non-stationary Markov chains. When applied to stationary Markov chains, the result clarifies the key difference between ergodicity and strong ergodicity and infinite stochastic matrices. More precisely, in Chapter III, Lemma III.1 shows that if $\{P_n\}$ is weakly ergodic and the limits of all diagonal entries of $P^{(m,m+t)}$ exist, then the limits of all entries of $P^{(m,m+t)}$ exist, and in Lemma III.2 we show that $\{P_n\}$ is weakly ergodic and $P_{jj}^{(m,m+t)}$ converges to $\pi_j \geqq 0$ for all j and m with $\sum_{j=1}^{\infty} \pi_j = 1$ if and only if $\{P_n\}$ is strongly ergodic. With the remark of Lemma III.2 we use Lemma III.1 and III.2 to prove the main theorem (Theorem III.3) of Chapter III, which characterizes uniform strong ergodicity using mean visit times. As a corollary (Corollary III.4) of the Theorem III.3 we have that a necessary and sufficient condition for a stationary chain to be strongly ergodic is that, for some aperiodic state j, the mean visit time to state j is finite for all starting vectors.

Some applied problems lead naturally to a systematic alteration of the transition matrices in order to reflect processes that are difficult or impossible to incorporate in the transition matrices of a Markov chains. In Chapter IV we consider the modification of finite non-homogeneous

Markov chains with at least one absorbing state. We apply to such chains the concept of normalization. (A sequence of substochastic matrices is said to be normalized if each matrix is right-multiplied by a certain positive diagonal matrix.) In section A, a scalar normalization is applied to wildlife migrations in order to maintain certain population densities or to prevent the extinction of the population in the total region or in the separate regions. In section B, we use minimal information (assume the existence of the bounds of the entries of the matrices of the sequence) to achieve bounds on wildlife population in the total area in a given time period or in the long run. In section D, the discussion of normalization suggests that the ergodic theory in Isaacson and Madsen (1974) can be pushed through for diagonal normalization. In section F, we describe a growth stock model in forestry. If a forest is not yielding periodic harvest equal to its periodic growth, then it is desirable to manage the forest by a long range plan in order to achieve sustained yield. To illustrate some of our theories we consider a very idealized approach which emphasizes a sustained growth stock and our goal is to stabilize the total growth stock.

II. THE RATE OF CONVERGENCE OF CERTAIN NON-HOMOGENEOUS MARKOV CHAINS

It is known that if the chain generated by a stochastic matrix, P, is weakly ergodic then the chain generated by P is strongly ergodic. If we view this chain as a "non-homogeneous" chain with $P_n = P$ for all n, then we have a trivial example where $P_n \to P$ and P is strongly ergodic implies that the non-stationary chain is strongly ergodic. We might wonder if the same result holds when the convergence of P_n to P is non-trivial. Mott (1957) proved that this result is true in the case of finite Markov chains. It has since been shown that the result is in fact true for a general state space [Bowerman, David and Isaacson (1975)].

Suppose we know that P is strongly ergodic and we know the rate of convergence of P_n to P, then it is natural to ask how fast the product $P_{m+1}P_{m+2}\cdots P_{m+n}$ converges. Will the rate of convergence be the same as that at which $P_n \to P$? In this chapter we consider the rate of convergence of $P_{m+1}P_{m+2}\cdots P_{m+n}$ to a row-constant matrix in terms of the rate of convergence of P_n to P.

Let $\{x_n\}_{n=1}^{\infty}$ be a non-homogeneous Markov chain with transition matrix $\{P_n\}_{n=1}^{\infty}$. Assume that P_n converges to a stochastic matrix P and that P is strongly ergodic (that is, P^n converges to a row-constant stochastic matrix Q). It is known [Bowerman, David and Isaacson (1975)] that in this case the product $P_{m+1}P_{m+2}\cdots P_{m+n}$ converges uniformly in m to the row-constant matrix Q. We now proceed to determine the rate of convergence of this product in terms of the rate of convergence of P_n to P.

The following lemmas will be used repeatedly in the proof of Theorem II.7. (See Paz, 1971 for proofs.)

<u>Lemma II.1</u>: If P and Q are stochastic matrices then

$$\delta(PQ) \leq \delta(P)\delta(Q).$$

<u>Lemma II.2</u>: If P is a stochastic matrix and if R is any real matrix such that

$$\sum_{k=1}^{\infty} r_{ik} = 0 \quad \text{for all } i, \quad \text{and} \quad \|R\| < \infty,$$

then,

$$\|RP\| \leq \|R\|\delta(P).$$

Lemma II.3: For real matrices A, B

$$\|AB\| \leq \|A\| \cdot \|B\|,$$

and
$$\|A+B\| \leq \|A\| + \|B\|.$$

Before considering the rate of convergence for non-stationary chains for which $P_n \to P$ we consider the special case where $P_n \equiv P$.

Lemma II.4: If P is the transition matrix for a homogeneous Markov chain and if P is strongly ergodic, then there exist constants c and $\beta\,(0 < \beta < 1)$ such that

$$\|P^n - Q\| \leq c\,\beta^n, \quad n = 1,2,3,\cdots.$$

Proof: Since P is strongly ergodic, it follows that $QP^n = Q$ for $n = 1,2,3,\cdots$ and there exist a positive integer N and a number d such that either

i) $0 < \delta(P^N) = d < 1.$

or

ii) $\delta(P^n) = 0$ for all n.

In case ii) P is itself a row constant matrix and the lemma is trivial. Hence we will just discuss case i).

Let $n \geq N$ and set $n = N \cdot q + r$ where $q = \left[\frac{n}{N}\right]$ and $0 \leq r < N$. Then,

$$\begin{aligned}
\|P^n - Q\| &= \|P^{r+Nq} - Q\| \\
&= \|P^{r+Nq} - QP^{Nq}\| \\
&\leq \|P^r - Q\|\delta(P^{Nq}) \\
&\leq 2(\delta(P^N))^q \\
&= 2d^q \\
&\leq \frac{2}{d}\left(d^{\frac{1}{N}}\right)^n.
\end{aligned}$$

Since $0 < d < 1$ and $N \geq 1$, we have $0 < d^{\frac{1}{N}} < 1$. Set

$\beta = d^{\frac{1}{N}}$ and $c = \frac{2}{d}$. Then $\|P^n - Q\| \leq c\beta^n$ for $n \geq N$.

Finally, if $n < N$ then

$\|P^n - Q\| < 2 = \frac{2}{\beta^n}\beta^n < \frac{2}{\beta^N}\beta^n = \frac{2}{d}\beta^n = c\beta^n$. Therefore

$$\|P^n - Q\| \leq c\beta^n \quad \text{for all} \quad n \geq 1.$$

Remark: If P is a finite transition matrix for a homogeneous Markov chain and if P is strongly ergodic then one can use the algebraic approach to find how fast P^n converges to the row-constant matrix Q. In particular if the eigenvalues of P are $\{\lambda_1, \lambda_2, \cdots \lambda_\ell, 1\}$ with $\max\limits_{1 \leq i \leq \ell} |\lambda_i| < 1$, then P^n converges to Q at a geometric rate as $n \to \infty$. That is $\|P^n - Q\| \leq k \max\limits_{1 \leq i \leq \ell} |\lambda_i|^n$ for all $n \geq 1$, where k is a constant. It can be shown that the rate of convergence obtained using the algebraic approach is always as fast as that obtained using the delta-coefficient in Lemma II.4. However, the delta-coefficient approach is often much easier to apply since it is easier to find $\delta(P)$ than the eigenvalues of P.

The next lemma uses the assumption $P_n \to P$ to show that the tail of $\{P_n\}$ has properties similar to that of P.

Lemma II.5: Let $\{P_n\}$ be the transition matrices for a non-homogeneous Markov chain and suppose that

$$\lim_{n\to\infty}\|P_n - P\| = 0 \quad \text{for some stochastic matrix } P.$$

i) If $\delta(P) < 1$, then there exist a positive integer N and a number β' such that

$$0 \leq \delta(P_n) \leq \beta' < 1 \quad \text{for} \quad n \geq N.$$

ii) If $\delta(P^M) < 1$ for some M, then there exist a positive integer T and a number $\nu' < 1$ such that

$$\delta(P^{(t,t+M)}) < \nu' \quad \text{for} \quad t \geq T.$$

Proof: i) It is known [Isaacson and Madsen (1976)] that

$$|\delta(P_n) - \delta(P)| \leq \|P_n - P\|.$$

Since $\delta(P) < 1$, we may set $1-\delta(P) = t > 0$. By assumption there exists N such that if $m \geq N$ then

$$\|P_m - P\| < \frac{t}{2} \text{ which implies that}$$

$$|\delta(p_m) - \delta(P)| < \frac{t}{2}.$$

Hence we have

$$0 \leq \delta(P_m) < \delta(P) + \frac{t}{2} < \delta(P) + t = 1.$$

Set $\beta' = \delta(P) + \frac{t}{2}$ and part i) follows.

ii) Since $P_n \to P$ it follows that $P_{\ell+1} \cdot \ldots \cdot P_{\ell+M} \to P^M$. [Isaacson and Madsen (1976)].

Hence using the same proof as in i) it follows that ii) holds.

<u>Lemma II.6</u>: Let $P_1, P_2, \cdots, P_n$ and P be stochastic matrices. Then,

$$\|P_1 P_2 \cdots P_n - P^n\| \leq \sum_{i=1}^{N} \|P_i - P\|.$$

Proof: By Lemma II.3, we have

$$\|P_1P_2\cdots P_n - P^n\| \leq \|P_1P_2\cdots P_n - P^{n-1}P_n\| + \|P^{n-1}P_n - P^n\|$$

$$\leq \|P_1P_2\cdots P_{n-1} - P^{n-1}\| + \|P_n - P\|$$

$$\vdots$$

$$\leq \|P_1 - P\| + \|P_2 - P\| + \cdots + \|P_n - P\|$$

$$= \sum_{i=1}^{n} \|P_i - P\|.$$

<u>Note</u>: By Lemma II.2 and Lemma II.3 it can be shown that

$$\|P_1P_2\cdots P_n - P^n\| \leq \|P_1P_2\cdots P_n - P^{n-1}P_n\| + \|P^{n-1}P_n - P^n\|$$

$$\leq \|P_1P_2\cdots P_{n-1} - P^{n-1}\|\delta(P_n) + \|P_n - P\|$$

$$\leq (\|P_1P_2\cdots P_{n-1} - P^{n-2}P_{n-1}\| + \|P^{n-2}P_{n-1} - P^{n-1}\|)\delta(P_n) + \|P_n - P\|$$

$$\leq \|P_1P_2\cdots P_{n-2} - P^{n-2}\|\delta(P_{n-1})\delta(P_n)$$

$$+ \|P_{n-1} - P\|\delta(P_n) + \|P_n - P\|$$

$$\vdots$$

$$\leq \sum_{i=1}^{n-1} [\|P_i - P\|\delta(P_{i+1})\delta(P_{i+2})\cdots\delta(P_n)]$$

$$+ \|P_n - P\|.$$

Since $\delta(P_k) \leq 1$ the second inequality is a better upper bound for $\|P_1P_2\cdots P_n - P^n\|$ than that in Lemma II.6. In Theorem II.7 these two inequalities will be used.

In the following theorem, we are going to find the rate of convergence of $P^{(t,t+n)}$ to Q in terms of the rate of convergence of P_n to P where P is strongly ergodic. The rate at which P_n converges to P will be specified by a monotone function $g(n)$. (Note that all functions of the form n^k or ℓ^n with $k > 0$ and $\ell \geq 1$ are candidates for $g(n)$.)

<u>Theorem II.7</u>: Let $\lim_{n\to\infty}\|P_n - P\| = 0$ where P is strongly ergodic. (By Lemma II.4 we have that $\|P^n - Q\| \leq c\beta^n$ for some $\beta(0 < \beta < 1)$ and constant c.) Let $g(n)$ be a monotone increasing function from the positive integers into the positive real numbers. If

$$\lim_{n\to\infty} g(2n)\|P_n - P\| = 0,$$

then not only is $\{P_n\}$ uniformly strongly ergodic [Bowerman, David and Isaacson (1975)], but

$$\lim_{n\to\infty} \sup_{t\geq 0}\{\min(\lambda^n, g(n))\|P^{(t,t+n)} - Q\|\} = 0,$$

where $1 < \lambda < \sqrt{1/\beta}$.

<u>Note</u>: When finding the inequality for $\|P^{(t,t+n)} - Q\|$ we need to use the fact that $g(n)$ is monotone increasing.

Proof: Since P is strongly ergodic, there exists a positive integer M such that $0 \leq \delta(P^M) < 1$. Hence by Lemma II.5 there exists a positive integer T and a number ν' such that

$$\delta(P^{(t,t+M)}) < \nu' < 1 \quad \text{for} \quad t \geq T.$$

From our hypothesis concerning $g(n)$, we have that given $\epsilon > 0$, there exists a positive integer $N = N(\epsilon)$ such that

$$(1) \qquad g(2n)\|P_n - P\| < \epsilon \quad \text{for} \quad n \geq N.$$

For notational convenience we assume that n is an even integer such that

$$\frac{n}{2} > \max(T,M,N).$$

(The proof is similar when n is odd.) Now, for $t = 0,1,2,\cdots$ we have

$$\|P^{(t,t+n)} - Q\| \leq \|P^{(t,t+n)} - P^{(t,t+\frac{n}{2})} P^{\frac{n}{2}}\| + \|P^{(t,t+\frac{n}{2})} P^{\frac{n}{2}} - Q\|$$

$$\text{(2)} \qquad \leq \|P^{(t+\frac{n}{2},t+n)} - P^{\frac{n}{2}}\| + \|P^{(t,t+\frac{n}{2})} P^{\frac{n}{2}} - P^{(t,t+\frac{n}{2})} Q\|$$

$$\leq \|P^{(t+\frac{n}{2},t+n)} - P^{\frac{n}{2}}\| + \|P^{\frac{n}{2}} - Q\|$$

The two terms on the right hand side of (2) will be considered separately. For the first term we have

$$\|P^{(t+\frac{n}{2},t+n)} - P^{\frac{n}{2}}\|$$

$$\leq \|P^{(t+\frac{n}{2},t+n)} - P^{\frac{n}{2}-M} P^{(t+n-M,t+n)}\| + \|P^{\frac{n}{2}-M} P^{(t+n-M,t+n)} - P^{\frac{n}{2}}\|$$

$$\leq \|P^{(t+\frac{n}{2},t+n-M)} - P^{\frac{n}{2}-M}\| \delta(P^{(t+n-M,t+n)})$$

$$+ \|P^{(t+n-M,t+n)} - P^{M}\|,$$

by Lemmas II.2 and II.3 we repeat the above procedure on $\|P^{(t+\frac{n}{2},t+n-M)} - P^{\frac{n}{2}-M}\|$ and again reduce the number of factors in $P^{(t+\frac{n}{2},t+n-M)}$ by M factors. Continuing this reduction using the fact that $\delta(P^{(t,t+M)}) < \nu'$ for $t > T$, we get

$$\|P^{(t+\frac{n}{2},t+n)} - P^{\frac{n}{2}}\| \leq \|P^{(t+\frac{n}{2},t+n-(q-1)M)} - P^{\frac{n}{2}-(q-1)M}\|(\nu')^{q-1}$$

$$+ \sum_{i=1}^{q-1} \|P^{(t+n-iM,t+n-(i-1)M)} - P^{M}\|(\nu')^{i-1}$$

where $q = \left[\frac{n}{2M}\right]$. Now set

$$B_n = g(n)\|P^{(t+\frac{n}{2},t+n-(q-1)M)} - P^{\frac{n}{2}-(q-1)M}\|.$$

By Lemma II.6, we have

$$B_n \le g(n)\left(\|P_{t+\frac{n}{2}+1} - P\| + \cdots + \|P_{t+n-(q-1)M} - P\|\right)$$

$$\le g(n+2t+2)\|P_{t+\frac{n}{2}+1} - P\| + g(n+2t+4)\|P_{t+\frac{n}{2}+2} - P\|$$

$$+ \cdots + g(2(t+n-(q-1)M))\|P_{t+n-(q-1)M} - P\|$$

Since $t+\frac{n}{2}+1 > \frac{n}{2} > \max(T,M,N)$ for $t \ge 0$, we have by (1) that

$$B_n < 2M\epsilon .$$

Similarly we can show that

$$g(n)\|P^{(t+n-iM,\,t+n-(i-1)M)} - P^M\| < M\epsilon, \quad i = 1,2,\cdots,q-1.$$

Thus

$$(3) \qquad g(n)\left\|P^{(t+\frac{n}{2},\,t+n)} - P^{\frac{n}{2}}\right\| < 2M\epsilon\left(\sum_{i=0}^{q-1} (\nu')^{i}\right)$$

$$< \left(\frac{2M}{1-\nu'}\right)\epsilon .$$

Hence the left hand side of (3) approaches zero uniformly in t.

The second term on the right hand side of (2) is less than $c\beta^{\frac{n}{2}}$ by Lemma II.4, where $0 < \beta < 1$. Hence if λ is a number such that $1 < \lambda < \sqrt{1/\beta}$, then

$$\min(g(n),\lambda^{n})\left\|P^{\frac{n}{2}} - Q\right\| \leq \lambda^{n}c\beta^{\frac{n}{2}}$$

$$= (\lambda^{2})^{\frac{n}{2}}c\beta^{\frac{n}{2}}$$

$$= c\left(\frac{\lambda^{2}}{1/\beta}\right)^{\frac{n}{2}} .$$

The latter expression clearly approaches zero as n approaches infinity. Therefore

$$\lim_{n\to\infty} \min(g(n),\lambda^n)\|P^{(t,t+n)} - Q\| = 0$$

uniformly in t where $t \geq 0$.

Remark: In Theorem II.7 if we assume that $\lim_{n\to\infty} g(kn)\|P_n - P\| = 0$ where k is a positive integer, then the result is true with $1 < \lambda < \left(\frac{1}{\beta}\right)^{\frac{k-1}{k}}$. The proof of this is obtained by changing the inequality (2) as follows

$$\|P^{(t,t+n)} - Q\| \leq \|P^{(t,t+n)} - P^{(t,t+\frac{n}{k})} P^{n-\frac{n}{k}}\|$$

$$+ \|P^{(t,t+\frac{n}{k})} P^{n-\frac{n}{k}} - Q\|,$$

and subsequently using the same technique as in the proof of the theorem.

If $\lim_{n\to\infty} g(kn)\|P_n - P\| = 0$ for all $k = 2,3,4,\cdots$, then the result is true with $1 < \lambda < \frac{1}{\beta}$.

__Example__: Let $P_n = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3}-\frac{1}{n}, & \frac{2}{3}+\frac{1}{n} \end{pmatrix} \to P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$, then P is strongly ergodic and $P^n \to Q = \begin{pmatrix} \frac{2}{5}, & \frac{3}{5} \\ \frac{2}{5}, & \frac{3}{5} \end{pmatrix}$.

$$\delta(P) = \frac{1}{2}\left(\left|\frac{1}{2}-\frac{1}{3}\right| + \left|\frac{1}{2}-\frac{2}{3}\right| \right) = \frac{1}{2}\left(\frac{1}{6}+\frac{1}{6}\right) = \frac{1}{6} < 1,$$

so we get

$$\begin{aligned} \|P^n - Q\| &= \|P^n - QP^n\| \\ &\leq \|I - Q\|\delta(P^n) \\ &\leq 2(\delta(P))^n \\ &= 2\left(\frac{1}{6}\right)^n \quad \text{for all } n. \end{aligned}$$

Also note that

$$\delta(P_n) = \frac{1}{2}\left(\left|\frac{1}{2} - \frac{1}{3} + \frac{1}{n}\right| + \left|\frac{1}{2} - \frac{2}{3} - \frac{1}{n}\right| \right)$$

$$= \frac{1}{2}\left(\frac{1}{6} + \frac{1}{n} + \frac{1}{6} + \frac{1}{n}\right)$$

$$= \frac{1}{6} + \frac{1}{n}$$

$$< \frac{1}{3} \quad \text{for} \quad n \geq 7,$$

and

$$P_n - P = \begin{pmatrix} 0 & 0 \\ -\frac{1}{n} & \frac{1}{n} \end{pmatrix} \quad \text{implies} \quad \|P_n - P\| = \frac{2}{n}.$$

If we let $g(n) = \sqrt{\frac{n}{2}}$, then

$$\lim_{n\to\infty} g(2n)\|P_n - P\| = \lim_{n\to\infty} \sqrt{\frac{2n}{2}} \quad \frac{2}{n} = 0.$$

Now without loss of generality assume that n is even, then for $n > 14$, we have

$$\|P^{(t,t+n)} - Q\| \le \|P^{(t+\frac{n}{2},t+n)} - P^{\frac{n}{2}}\| + \|P^{\frac{n}{2}} - Q\|$$

$$\le \sum_{i=1}^{n/2} \|P_{\frac{n}{2}+t+i} - P\| \left(\frac{1}{3}\right)^{\frac{n}{2}-i} + 2\left(\frac{1}{6}\right)^{\frac{n}{2}}.$$

Since $\delta(P) = \frac{1}{6}$, we set $\beta = \frac{1}{6}$ and let $\lambda = 2(1 < 2 < \sqrt{6})$, then for $n > 14$ and all t we have

$$\min(g(n), \lambda^n) \|P^{(t,t+n)} - Q\|$$

$$= \min\left(\sqrt{\frac{n}{2}}, 2^n\right) \|P^{(t,t+n)} - Q\|$$

$$= \sqrt{\frac{n}{2}} \|P^{(t,t+n)} - Q\|$$

$$\le \frac{\sqrt{n}}{\sqrt{2}} \left(\sum_{i=1}^{n/2} \|P_{\frac{n}{2}+t+i} - P\| \left(\frac{1}{3}\right)^{\frac{n}{2}-i} + 2\left(\frac{1}{6}\right)^{\frac{n}{2}} \right)$$

$$\le \frac{\sqrt{n}}{\sqrt{2}} \frac{2}{\frac{n}{2} + t + 1} \sum_{i=1}^{n/2} \left(\frac{1}{3}\right)^{\frac{n}{2}-i} + \frac{\sqrt{n}}{\sqrt{2}} (2) \frac{1}{\sqrt{6^n}}$$

$$\leq \frac{\sqrt{n}\sqrt{2}}{n/2} \quad \frac{1}{1-\frac{1}{3}} + \frac{2\sqrt{n}}{\sqrt{2}\sqrt{6^n}}$$

$$= \frac{3\sqrt{2}}{\sqrt{n}} + \frac{\sqrt{2}\sqrt{n}}{\sqrt{6^n}} .$$

Thus

$$\lim_{n\to\infty} \sup_{t\geq 0}\{\min(g(n),\lambda^n)\|P^{(t,t+n)} - Q\|\}$$

$$\leq \lim_{n\to\infty}\left(\frac{3\sqrt{2}}{\sqrt{n}} + \frac{\sqrt{2}\sqrt{n}}{\sqrt{6^n}}\right)$$

$$= 0.$$

Note that the rate of convergence was dictated by the rate at which P_n converged to P.

The next example will have $\min(g(n),\lambda^n) = \lambda^n$ so that the rate at which $P^n \to Q$ dictates the rate at which

$P^{(t,t+n)} \to Q$.

Example: Let $P_n = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} - \frac{1}{2^{n+1}}, & \frac{2}{3} + \frac{1}{2^{n+1}} \end{pmatrix} \to P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$. Then P is strongly ergodic and $P^n \to Q = \begin{pmatrix} \frac{2}{5}, & \frac{3}{5} \\ \frac{2}{5}, & \frac{3}{5} \end{pmatrix}$. Since $\delta(P) = \frac{1}{6} < 1$, $\|P^n - Q\| \leq 2\left(\frac{1}{6}\right)^n$ for all n. Also note that

$$\delta(P_n) = \frac{1}{2}\left(\left|\frac{1}{2} - \frac{1}{3} + \frac{1}{2^{n+1}}\right| + \left|\frac{1}{2} - \frac{3}{2} - \frac{1}{2^{n+1}}\right| \right)$$

$$= \frac{1}{2}\left(\frac{1}{6} + \frac{1}{2^{n+1}} + \frac{1}{6} + \frac{1}{2^{n+1}}\right)$$

$$= \frac{1}{6} + \frac{1}{2^{n+1}}$$

$$< \frac{1}{3} \quad \text{for} \quad n \geq 2,$$

and $\|P_n - P\| = \frac{2}{2^{n+1}} = \frac{1}{2^n}$. If we let $g(n) = 2^{\frac{n}{4}}$, then

$$\lim_{n\to\infty} g(2n)\|P_n - P\| = \lim_{n\to\infty} 2^{\frac{n}{2}}\left(\frac{1}{2^n}\right) = 0.$$

Now without loss of generality assume that n is even, then for $n \geq 4$, we have

$$\|P^{(t,t+n)} - Q\| \leq \|P^{(t+\frac{n}{2},t+n)} - P^{\frac{n}{2}}\| + \|P^{\frac{n}{2}} - Q\|$$

$$\leq \sum_{i=1}^{n/2} \|P_{\frac{n}{2}+t+i} - P\|\left(\frac{1}{3}\right)^{\frac{n}{2}-i} + 2\left(\frac{1}{6}\right)^{\frac{n}{2}}.$$

Since $\delta(P) = \frac{1}{6}$, we set $\beta = \frac{1}{6}$ and let $\lambda = 2^{\frac{1}{5}}$ $\left(1 < 2^{\frac{1}{5}} < \sqrt{6}\right)$, then for $n \geq 14$ and all t we have (λ can be chosen larger than 2 so again $g(n) \leq \lambda^n$.)

$$\min(g(n),\lambda^n)\,\|P^{(t,t+n)} - Q\|$$

$$= \min\left(2^{\frac{n}{4}}, 2^{\frac{n}{5}}\right)\|P^{(t,t+n)} - Q\|$$

$$= 2^{\frac{n}{5}}\|P^{(t,t+n)} - Q\|$$

$$\le 2^{\frac{n}{5}}\left(\sum_{i=1}^{n/2} \|P_{\frac{n}{2}+t+i} - P\|\left(\frac{1}{3}\right)^{\frac{n}{2}-i} + 2\left(\frac{1}{6}\right)^{\frac{n}{2}}\right)$$

$$\le 2^{\frac{n}{5}}\,\frac{1}{2^{\frac{n}{2}+t+1}}\sum_{i=1}^{n/2}\left(\frac{1}{3}\right)^{\frac{n}{2}-i} + \left(2^{\frac{n}{5}}\right)(2)\left(\frac{1}{6^{\frac{n}{2}}}\right)$$

$$\le \frac{1}{2^{\frac{3n}{10}+t+1}}\,\frac{1}{1-\frac{1}{3}} + (2)\left(\frac{2^2}{6^5}\right)^{\frac{n}{10}}.$$

Thus

$$\lim_{n\to\infty} \sup_{t\geq 0}\left\{\min(g(n),\lambda^n)\|P^{(t,t+n)} - Q\|\right\}$$

$$\leq \lim_{n\to\infty}\left(\frac{3}{2^{\frac{3n}{2}+t+2}} + 2\left(\frac{4}{6}\right)^{\frac{n}{10}}\right)$$

$$= 0.$$

Note: Different rates of convergence of P_n to P are used in the above two examples, with $\min(g(n),\lambda^n) = g(n)$ in the first example, and $\min(g(n),\lambda^n) = \lambda^n$ in the second example.

A particular type of non-homogeneous Markov chain recently discussed in the literature is the constant causative chain (see for example [Pullman and Styan (1973)]). A matrix C is said to be causative with respect to a matrix P_1 if P_1C^n is stochastic for $n \geq 0$. The matrix C is said to be causative when it is causative with respect to some matrix P_1. The following corollary has as a simple consequence the validty of a conjecture of [Lipstein (1965)] (see [Pullman and Styan (1973)] for an alternate proof).

<u>Corollary II.8</u>: Let $P_n = P_1 C^{n-1}$, where C is of infinite order, causative with respect to P_1 and bounded. If $\lim_{n\to\infty} C^n = L$, where L is a row-constant stochastic matrix, then there exist constants k and $\rho\ (0 < \rho < 1)$ such that $\|P^{(t,t+n)} - L\| \leq k\rho^n$ uniformly in t, $t \geq 0$.

Proof: Since $C^n \to L$, there exist a positive integer N and a number α such that

$$0 \leq \|C^n - L\| < \alpha < 1 \quad \text{for} \quad n \geq N,$$

and for C and L as given it can be shown that $C^n - L = (C - L)^n$ for all n [Pullman and Styan (1973)]. Using these facts we can show that $P_n \to L$ at a geometric rate. For this causative chain we have $P = Q = L$ so $\|P^n - Q\| \equiv 0$. Hence by the Theorem II.7 $\|P^{(t,t+n)} - L\| \to 0$ at a geometric rate uniformly in t.

III. ERGODICITY USING MEAN VISIT TIMES

For this chapter we need the following definitions and notation.

<u>Definition III.1</u>: For a stationary Markov chain with transition matrix P, we define the mean visit time from state i to state j as

$$m_{ij} = \sum_{k=1}^{\infty} k f_{ij}^{(k)}, \quad \text{where} \quad f_{ij}^{(k)} = \text{prob.}$$

{first visit j from i at step k}.

<u>Definition III.2</u>: For a non-stationary Markov chain $\{x_n\}$ with transition matrices $\{P_n\}$, we define

$$_s m_{ij} = \sum_{k=1}^{\infty} k h_{ij}^{(s,s+k)},$$

where $h_{ij}^{(s,s+k)}$ = prob. {first visit j from i at the time $s+k$ starting at time $s+1$}.

Definition III.3: P is C-strongly ergodic if

$$\left\| \frac{1}{n} \sum_{k=1}^{n} P^k - Q \right\| \to 0 \quad \text{as} \quad n \to \infty$$

where Q is a row-constant stochastic matrix. [Bowerman, David and Isaacson (1977)].

Pitman (1974) uses mean visit times to discuss the uniform rate of convergence of a Markov chain. Pitman also points out that an infinite ergodic chain may have infinite mean visit time for certain starting vectors. In this chapter we use the concept of mean visit times to give necessary and sufficient conditions for a non-stationary chain to be uniformly strongly ergodic. As a corollary of this theorem we have that a necessary and sufficient condition for a stationary chain to be weakly ergodic is that for some aperiodic state j, the mean visit time to state j is finite for all starting vectors.

Our first lemma shows that if $\{P_n\}$ is weakly ergodic and the limits of all diagonal entries of $P^{(m,m+t)}$ exist, then the limits of all entries of $P^{(m,m+t)}$ exist.

<u>Lemma III.1</u>: If $\{P_n\}$ is weakly ergodic and $\lim_{t\to\infty} p_{\ell\ell}^{(m,m+t)} = \pi_\ell$, then $\lim_{t\to\infty} p_{i\ell}^{(m,m+t)} = \pi_\ell$ for all i.

Proof: By the definition of weak ergodicity we have that for all m and all starting vectors μ and λ

$$\lim_{t\to\infty} \sup_{\mu,\lambda} \|\mu P^{(m,m+t)} - \lambda P^{(m,m+t)}\| = 0.$$

Let $e_i = (0,0,\cdots,1,0,\cdots)$ where the one is in the $i^{\underline{th}}$ coordinate. We have

$$\lim_{t\to\infty} \|e_i P^{(m,m+t)} - e_k P^{(m,m+t)}\| = 0$$

uniformly in i and k. That is, for given $\epsilon > 0$, there exists $T_1 = T_1(m,\epsilon)$ such that

$$\|e_i P^{(m,m+t)} - e_k P^{(m,m+t)}\| < \frac{\epsilon}{2} \text{ provided } t \geq T_1.$$

Therefore

$$|p_{ij}^{(m,m+t)} - p_{kj}^{(m,m+t)}| < \frac{\epsilon}{2}$$

for all i, j, k and $t \geq T_1$. Now $\lim_{t\to\infty} p_{\ell\ell}^{(m,m+t)} = \pi_\ell$ implies there exists $T_2 = T_2(m,\epsilon)$ such that

$$|p_{\ell\ell}^{(m,m+t)} - \pi_\ell| < \frac{\epsilon}{2} \quad \text{provided} \quad t \geq T_2.$$

Let $T = T_1 + T_2$. Then for all i,

$$\begin{aligned} |p_{i\ell}^{(m,m+t)} - \pi_\ell| &\leq |p_{i\ell}^{(m,m+t)} - p_{\ell\ell}^{(m,m+t)}| + |p_{\ell\ell}^{(m,m+t)} - \pi_\ell| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon \quad \text{provided} \quad t \geq T. \end{aligned}$$

Hence $\lim_{t\to\infty} p_{i\ell}^{(m,m+t)} = \pi_\ell$ for all i.

<u>Note</u>: The above proof shows that if $\{P_n\}$ is weakly ergodic and if $\lim_{t\to\infty} p_{jj}^{(m,m+t)} = \pi_j$ for all j, then for all j, $\lim_{t\to\infty} p_{ij}^{(m,m+t)} = \pi_j$ uniformly in i.

<u>Lemma III.2</u>: $\{P_n\}$ is weakly ergodic and $p_{jj}^{(m,m+t)} \to \pi_j \geq 0$ for all j and m with $\sum_{j=1}^{\infty} \pi_j = 1$ if and only if $\{P_n\}$ is strongly ergodic.

Proof: Strong ergodicity implies the other three conditions by definition so we only prove the converse.

Since $\sum_{j=1}^{\infty} \pi_j = 1$, we have, given that $\epsilon > 0$, there exists N such that $\sum_{j=n}^{\infty} \pi_j < \epsilon$ provided $n \geq N$.

By the note of Lemma III.1., there exists $T = T(\epsilon,m,N)$ such that for all i and $j \leq N - 1$,

$$|p_{ij}^{(m,m+t)} - \pi_j| < \frac{\epsilon}{N-1} \quad \text{provided} \quad t \geq T.$$

Thus for all i,

$$\left|\sum_{j=N}^{\infty} p_{ij}^{(m,m+t)} - \sum_{j=N}^{\infty} \pi_j\right|$$

$$= \left|1 - \sum_{j=1}^{N-1} p_{ij}^{(m,m+t)} - \left(1 - \sum_{j=1}^{N-1} \pi_j\right)\right|$$

$$\leq \sum_{j=1}^{N-1} \left|p_{ij}^{(m,m+t)} - \pi_j\right| < (N-1) \cdot \frac{\epsilon}{N-1} = \epsilon$$

provided $t \geq T$. Therefore, for all i, m and $t \geq T$,

$$\sum_{j=1}^{\infty} \left|p_{ij}^{(m,m+t)} - \pi_j\right| \leq \sum_{j=1}^{N-1} \left|p_{ij}^{(m,m+t)} - \pi_j\right| + \sum_{j=N}^{\infty} p_{ij}^{(m,m+t)} + \sum_{j=N}^{\infty} \pi_j$$

$$\leq \epsilon + 2\epsilon + \epsilon$$

$$= 4\epsilon .$$

Thus, for all m,

$$\sup_{i} \sum_{j=1}^{\infty} |p_{ij}^{(m,m+t)} - \pi_j| \leq 4\epsilon, \quad \text{provided} \quad t \geq T,$$

and thus for all m, $\|P^{(m,m+t)} - Q\| \leq 4\epsilon$ provided $t \geq T$, where

$$Q = \begin{pmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \cdot (\pi_1, \pi_2, \pi_3, \cdots).$$

Hence $\{P_n\}$ is strongly ergodic.

Remark: The above lemma is true if we change weakly ergodic and strongly ergodic to uniformly weakly ergodic and uniformly strongly ergodic respectively, and assume that $p_{jj}^{(m,m+t)} \to \pi_j$ uniformly in m for all j.

The need for the assumption $\sum_{j=1}^{\infty} \pi_j = 1$ in Lemma III.2 is demonstrated by the following example:

Example: If

$$P_n = \begin{pmatrix} \frac{1}{n} & 1-\frac{1}{n} & 0 & 0 & \cdot & \cdot & \cdot \\ \frac{1}{n} & 0 & 1-\frac{1}{n} & 0 & \cdot & \cdot & \cdot \\ \frac{1}{n} & 0 & 0 & 1-\frac{1}{n} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \\ \cdot & \cdot & \cdot & \cdot & & & \cdot \end{pmatrix},$$

Then $\pi_j = 0$ for all j. Therefore $\{P_n\}$ is not strongly ergodic. However, $\delta(P_n) = 1 - \frac{1}{n}$, so $\{P_n\}$ is weakly ergodic. (See Dobrushin, (1956).)

We now gave a characterization of uniform strong ergodicity using mean visit times.

Theorem III.3: $\{P_n\}$ is uniformly strongly ergodic if and only if for every j, $\lim_{t\to\infty} p_{jj}^{(m,m+t)} = \pi_j \geq 0$ uniformly in m, $\sum_i \pi_i = 1$ and for j^* with $\pi_{j^*} > 0$, it follows that $\sup_{s,i}({}_s m_{ij^*}) < \infty$. In fact, uniform strong ergodicity implies

$\sup_{s,i}({}_s m_{ij}) < \infty$ for all j with $\pi_j > 0$.

Proof: Assume that $\{P_n\}$ is uniformly strongly ergodic. By definition we know that for every j,

$\lim_{t\to\infty} p_{jj}^{(m,m+t)} = \pi_j \geq 0$ uniformly in m and also $\sum_{i=1}^{\infty} \pi_i = 1$.

Thus we only need to show that $\sup_{s,i}({}_s m_{ij^*}) < \infty$, where $\pi_{j^*} > 0$.

Let Q be the row-constant stochastic matrix for which

$$\lim_{t\to\infty} \| P^{(m,m+t)} - Q \| = 0 \text{ uniformly in } m.$$

(Denote the common row of Q by $(\pi_1, \pi_2, \pi_3, \cdots)$.) That is, there exists $T = T(\pi_{j^*})$ such that $\| P^{(m,m+t)} - Q \| < \frac{\pi_{j^*}}{2}$ provided $t \geq T$, or in other words

$$\sup_i \sum_{\ell} |p_{i\ell}^{(m,m+t)} - \pi_\ell| < \frac{\pi_{j^*}}{2} \quad \text{for} \quad t \geq T.$$

In particular

$$|p_{ij*}^{(m,m+t)} - \pi_{j*}| < \frac{\pi_{j*}}{2} \quad \text{for all} \quad i,\ m \quad \text{and} \quad t \geq T.$$

Thus

$$\inf_i p_{ij*}^{(m,m+t)} \geq \frac{\pi_{j*}}{2} \quad \text{for all} \quad m \quad \text{and} \quad t \geq T.$$

Now for $r = 0,1,2,\cdots,T-1$

$$h_{ij*}^{(m,m+2T+r)} = \sum_{k \neq j*} p_{ik}^{(m,m+T)} h_{kj*}^{(m+T,m+2T+r)}$$

$$\leq \sum_{k \neq j*} p_{ik}^{(m,m+T)}$$

$$\leq 1 - \frac{\pi_{j*}}{2} \quad \text{for all} \quad m \quad \text{and} \quad i,$$

and

$$h_{ij*}^{(m,m+3T+r)} \leq \sum_{k\neq j*}' p_{ik}^{(m,m+T)} h_{kj*}^{(m+T,m+3T+r)}$$

$$\leq \left(1 - \frac{\pi_{j*}}{2}\right) \sum_{k\neq j*} p_{ik}^{(m,m+T)}$$

$$\leq \left(1 - \frac{\pi_{j*}}{2}\right)\left(1 - \frac{\pi_{j*}}{2}\right)$$

$$= \left(1 - \frac{\pi_{j*}}{2}\right)^2 \quad \text{for all } m \text{ and } i.$$

By induction we have for $r = 0,1,2,\cdots,T-1$ and $m = 2,3,4\cdots$

$$h_{ij*}^{(m,m+nT+r)} \leq \left(1 - \frac{\pi_{j*}}{2}\right)^{n-1} \quad \text{for all } m \text{ and } i.$$

Therefore for all i and s,

$$({}_s m_{ij*}) = \sum_{k=1}^{\infty} k h_{ij*}^{(s,s+k)}$$

$$\leq \sum_{k=1}^{2T-1} kh_{ij*}^{(s,s+k)} + \sum_{k=2T}^{\infty} kh_{ij*}^{(s,s+k)}$$

$$= \sum_{k=1}^{2T-1} kh_{ij*}^{(s,s+k)} + \sum_{k=2T}^{3T-1} kh_{ij*}^{(s,s+k)}$$

$$+ \sum_{k=3T}^{4T-1} kh_{ij*}^{(s,s+k)} + \cdots$$

$$\leq \sum_{k=1}^{2T-1} kh_{ij*}^{(s,s+k)} + (3T)(T)\left(1 - \frac{\pi_{j*}}{2}\right)$$

$$+ (4T)(T)\left(1 - \frac{\pi_{j*}}{2}\right)^2 + \cdots$$

$$\leq \sum_{k=1}^{2T-1} k + \sum_{n=3}^{\infty} nT^2\left(1 - \frac{\pi_{j*}}{2}\right)^{n-2}$$

$= B < \infty$ independently of i and s.

Thus

$$\sup_{s,i}({}_s m_{ij^*}) < \infty.$$

For the converse, suppose $\{P_n\}$ is not uniformly strongly ergodic, then by the remark of Lemma III.2 we know that $\{P_n\}$ is not uniformly weakly ergodic, therefore given t there exists $m = M$ such that

$$\delta(P^{(M,M+t)}) > 1 - \frac{\pi_{j^*}}{4} .$$

That is, for every t, there exists i_t such that

$$p_{i_t,j^*}^{(M,M+t)} < \frac{\pi_{j^*}}{4} . \tag{1}$$

By hypothesis

$$\lim_{t\to\infty} p_{j^*j^*}^{(m,m+t)} = \pi_{j^*} > 0 \quad \text{uniformly in } m,$$

so there exists T such that

$$p_{j^*j^*}^{(M,M+t)} > \frac{\pi_{j^*}}{2} > 0 \quad \text{provided} \quad t \geq T.$$

Let $t \geq 2T$, then

$$p_{i_t,j^*}^{(M,M+t)} = \sum_{k=1}^{t-1} h_{i_t j^*}^{(M,M+k)} p_{j^*j^*}^{(M+k,M+t)}$$

$$\geq \sum_{k=1}^{[t/2]} h_{i_t j^*}^{(M,M+k)} p_{j^*j^*}^{(M+k,M+t)}$$

$$> \frac{\pi_{j^*}}{2} \sum_{k=1}^{[t/2]} h_{i_t j^*}^{(M,M+k)} .$$

By (1) we have

$$\frac{\pi_{j^*}}{4} > \frac{\pi_{j^*}}{2} \sum_{k=1}^{[t/2]} h_{i_t j^*}^{(M,M+k)} .$$

That is,

$$\frac{1}{2} > \sum_{k=1}^{[t/2]} h_{i_t j^*}^{(M,M+k)} \quad \text{for} \quad t \geq 2T.$$

Hence

$$\left({}_M m_{i_t j^*}\right) = \sum_{k=1}^{\infty} k h_{i_t j^*}^{(M,M+k)}$$

$$> ([t/2]+1) \sum_{k=[t/2]+1}^{\infty} h_{i_t j^*}^{(M,M+k)}$$

$$> ([t/2]+1)(1-1/2) \left(\text{Since} \quad \sum_{k=1}^{\infty} h_{i_t j^*}^{(M,M+k)} = 1 \quad \text{when} \quad m_{i_t j^*} < \infty\right)$$

$$\geq \frac{t+1}{4} \quad \text{for} \quad t \geq 2T.$$

Hence $\sup_{s,i}({}_s m_{ij^*}) = \infty$ and this contradicts the hypothesis.

Therefore $\{P_n\}$ is uniformly strongly ergodic.

Using the proof of the above theorem the following corollary can be proved easily by changing $h_{ij}^{(m,m+k)}$ to $f_{ij}^{(k)}$.

<u>Corollary III.4</u>: The stationary chain P is weakly ergodic if and only if $\sup_i m_{ij*} < \infty$ for some $j*$ where $j*$ is an aperiodic positive persistent state. In fact, weak ergodicity implies that $\sup_i m_{ij} < \infty$ for all j with $\pi_j > 0$.

Proof: Using the fact that for a stationary Markov chain weak ergodicity is equivalent to uniform strong ergodicity, the corollary follows from Theorem III.3.

<u>Corollary III.5</u>: If $\sup_i m_{ij*} < \infty$ for some $j*$, then the stationary chain P is C-strongly ergodic.

Proof: Recall that P is C-strongly ergodic if

$$\left\|\frac{1}{n}\sum_{k=1}^{\infty} P^k - Q\right\| \to 0 \quad \text{as} \quad n \to \infty$$

where Q is a row constant stochastic matrix [see Bowerman, David and Isaacson (1977)]. Let d be the period of the persistent state j^*. If $d = 1$, then by Corollary III.4 we know that P is weakly ergodic hence P is C-strongly ergodic. If $d \neq 1$, then for notational convenience we will assume that P is irreducible and of the form

$$P = \begin{pmatrix} 0 & P_1 & 0 & \cdots & 0 \\ 0 & 0 & P_2 & \cdots & 0 \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ 0 & 0 & 0 & \cdots & P_{d-1} \\ P_d & 0 & 0 & \cdots & 0 \end{pmatrix}$$

In this case

$$P^d = \begin{pmatrix} (P_1 \cdots P_d) & 0 & 0 & \cdots & 0 \\ 0 & (P_2 P_3 \cdots P_d P_1) & 0 & \cdots & 0 \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ 0 & 0 & 0 & \cdots & (P_d P_1 \cdots P_{d-1}) \end{pmatrix} .$$

Each of the matrices $R_k = P_k P_{k+1} \cdots P_d P_1 \cdots P_{k-1}$ is aperiodic for $k = 1,2,\cdots d$. It can be shown [see Bowerman, David and Isaacson (1977)] that if one of these matrices, R_k, is weakly ergodic then they all are. By Corollary III.4 the R_k associated with the state j^* is weakly ergodic so the subsequence P^{nd+k} converges as $n \to \infty$ for $k = 1,2,\cdots d$. Hence P is C-strongly ergodic.

Applications: It is well-known that in a stationary Markov chain, if $m_{ii} = \infty$, then state i is null persistent (note if, state i is transient we consider m_{ii} undefined) and so are all states that intercommunicate with i. Let C_i denote all states that intercommunicate with i.

Similarly if $m_{ii} < \infty$, then state i is positive persistent and so are all states that intercommunicate with i. In this case the class C_i is C-ergodic. $\left[\text{that is, } \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{k} = \pi_j\right]$ and if i is aperiodic then the class C_i is ergodic.

We now consider state i with the property that $\sup_{j \in C_i} m_{ji} < \infty$. Such states will be called uniformly positive persistent on C_i. By Corollary III.5 we know that if state i is uniformly positive persistent on C_i, then the Markov chain with state space C_i is C-strongly ergodic. Similarly from Corollary III.4, if state i is aperiodic then the Markov chain with state space C_i is strongly ergodic. It also follows from Corollary III.4 that if state i is uniformly positive persistent on C_i then so are all states in C_i. So again we get the result that states that intercommunicate are of the same type.

Remark: During the works of this chapter one optional fact about the characteristic of a weakly ergodic (or strongly ergodic) stationary Markov chain was found.

Theorem III.6: P is weakly ergodic if and only if $\inf_i p_{ij}^{(n)} > 0$ for some j and n.

Proof: Let P be weakly ergodic (i.e. strongly ergodic) and $P^n \to Q$ where Q is row-constant with each row $(\pi_1, \pi_2, \pi_3, \cdots)$, then there exists j such that $\pi_j > 0$.

Let $\epsilon = \frac{\pi_j}{2}$ then there exists N such that

$$\|P^n - Q\| < \frac{\pi_j}{2} \text{ provided } n \geq N.$$

By definition $\sup_i \sum_j |p_{ij}^{(n)} - \pi_j| < \frac{\pi_j}{2}$, i.e.

$|p_{ij}^{(n)} - \pi_j| < \frac{\pi_j}{2}$ for all i, thus $p_{ij}^{(n)} > \frac{\pi_j}{2}$ for all i.

Therefore $\inf_i p_{ij}^{(n)} \geq \frac{\pi_j}{2} > 0$.

For the converse, since $\inf_i p_{ij}^{(n)} > 0$ for some j and n, implies $\delta(P^n) < 1$, it follows that P is weakly ergodic.

IV. NORMALIZED FINITE NON-HOMOGENEOUS CHAINS AND APPLICATIONS

In this chapter we consider modifications of finite non-homogeneous Markov chains with at least one absorbing state. We apply to such chains the following concept of normalization:

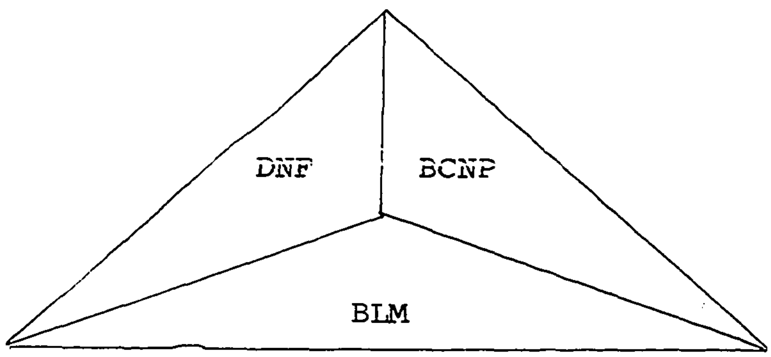
A sequence of substochastic matrices is said to be normalized if each matrix is right-multiplied by a positive diagonal matrix, generally a function of time, for the purpose of controlling the short or long range behavior of the product of the modified matrices. When it is desirable to distinguish the case where the diagonal matrix is scalar, we speak of scalar normalization.

In the applications discussed below, the normalization usually reflects "growth rates", in which case the diagonal matrix of the above definition is called a growth matrix and its entries are greater than or equal to 1.

It should be emphasized that the topics in this chapter are suggestive and not explored in depth.

A. The Wildlife Migration Model

As a first example of normalization we generalize the homogeneous example in the note by Kabak (1970), where a scalar normalization is applied to wildlife migrations. Consider 3 contiguous regions of public land in each of which the same species of wildlife exists or is stocked, and for which migration is possible from one region to another. It is desired to maintain certain population densities or to prevent the extinction of the population in the total region or in the separate regions. For instance, consider the particular case represented schematically by



where the three regions among which migrations occur are Bryce Canyon National Park where hunting is forbidden, Dixie National Forest where hunting is controlled, and Bureau of Land Management territory where hunting is also controlled.

We may idealize this situation as a normalized Markov process as follows:

i) The time is measured in years and transitions from one state to another are completed at end of the hunting season each fall. The states for an animal are:

(0) Death (this may be subdivided if desired according to cause, but will be taken as a cumulative figure from initial time);

(1) Living in BCNP;

(2) Living in DNF;

(3) Living in BLM.

The transition probabilities $p_t(i,j)$ are assumed dependent on i, j and t because the animal migration and mortality rates are affected unevenly in time by at least the following factors: forage, terrain, weather, predators, hunting, disease, noise, recreation, timbering and mining. Some of these are administratively manageable; others are predictable only within wide limits. In this three region case, with one death state, we introduce an absorbing state chain

$$P_t = \left(\begin{array}{c|c} 1 & 0 \quad 0 \quad 0 \\ \hline Q_t & A_t \end{array}\right)$$

where A_t is a "positive" substochastic 3×3 matrix. (Note: We use "positive" to mean every entry is positive.)

ii) The effects of birth rates, restocking, and any other positive influences on population density not taken into account by P_t are reflected in a normalization by a growth matrix

$G_t = \text{diag}(\gamma_t(0), \gamma_t(1), \cdots, \gamma_t(3)), \gamma_t(0) = 1, 1 \leq \gamma_t(i) \leq L.$

Let $H_t = \text{diag}(r_t(1), \cdots, r_t(3))$.

The operation during the transition period is therefore the positive matrix $A_t H_t$.

iii) Given an initial population distribution

$$\zeta = [\zeta_{01}, \zeta_{02}, \zeta_{03}],$$

let $$\zeta_N = \zeta \prod_{t=1}^{N} A_t H_t .$$

It is desired to predict or control ζ_N for certain time intervals or as $N \to \infty$ by using whatever freedom is available in G_t and A_t. Hunting fees, enforcement policies, management of grazing and timber and recreation are some examples of management tools which sensibly affect the behavior of resident wildlife. Both A_t and G_t are thus susceptible to significant manipulation.

In general, suppose we have $n-1$ regions and that annual migration rates from each region to the other regions is recorded and a record is kept of dead animals. As indicated above, we use the regions as $n-1$ possible states for live animals with death being the remaining state or $0^{\underline{th}}$ state.

Let P_t denote the transition matrix for the states at time t, so

$$P_t = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \cdot \\ \cdot \\ \cdot \\ n-1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1,2,\cdots,n-1 \end{array} \\ \left(\begin{array}{c|c} 1 & 0 \\ \hline Q_t & A_t \end{array}\right) \end{array}$$

where A_t is an $(n-1) \times (n-1)$ matrix and A_t is positive, state 0 is the death state.

A minimal positive mortality rate is built into the system by hunting, recreation, and herd controls; in addition, a variety of terrain and altitude causes variations in forage and breeding conditions. Hence we assume that

$$p_t(1,0) \geq \alpha,\ p_t(2,0) \geq \alpha, \cdots, p_t(n-1,0) \geq \alpha \quad \text{for all} \quad t$$

where α is a constant such that

$$0 < \alpha < 1.$$

Obviously, a vast amount of data and administrative discretion is required to implement our model. Such information and discretion is currently unavailable, so in the next two sections we analyze the model under the assumption of certain minimal information.

B. Wildlife Model: Minimal Information with Scalar Normalization

Suppose past experience suggests that it is possible to place bounds on $p_t(i,j)$ of the type

$$(*) \qquad 0 < \underline{a}(i,j) \leq p_t(i,j) \leq \bar{a}(i,j) < 1,$$

$$i,j = 1,2,3,\cdots,n-1,$$

$$t \geq 0,$$

where

$$\bar{a}(i,j) = \sup_t p_t(i,j)$$

$$\underline{a}(i,j) = \inf_t p_t(i,j).$$

We shall assume A_t is subject to (*). The set $\{A_t \mid t \geq 0\}$ contains convergent subsequences (because of (*)). If we have no way of determining these, then we may resort to the use of $\underline{A} = (\underline{a}(i,j))$ and $\bar{A} = (\bar{a}(i,j))$ as an approximating device.

Note that with the above assumptions, $\bar{A}$ may have row sums greater than 1, but $0 < \underline{A} \leq \bar{A}$ so $\underline{A}$ and $\bar{A}$ are positive matrices. In what follows it is convenient, although not always necessary, to assume that $\bar{A}$ is properly substochastic. Also note that any limit point of $\{A_t \mid t \geq 0\}$ satisfies (*).

Let us then proceed as though it is realistic to assume the above knowledge of bounds on A_t, that is, we know $\underline{A}$ and $\bar{A}$ such that

$$0 < \underline{A} = \inf_t(a_t(i,j)) \leq A_t \leq \sup_t(a_t(i,j)) = \bar{A} < A,$$

where $\bar{A}$ and $\underline{A}$ are substochastic matrices, and A is a stochastic matrix. Furthermore we assume that the growth rates are the same in each region, that is, the normalization is scalar. Hence we use a scalar growth matrix,

$$G_t = \begin{pmatrix} 1 & 0 \\ 0 & r_t I_{n-1} \end{pmatrix}, \quad r_t \geq 1,$$

where I_{n-1} is an $(n-1) \times (n-1)$ identity matrix.

Now $\prod_{t=1}^{N} A_t \to 0$ as $N \to \infty$. (Since $p_t(i,0) \geq \alpha > 0$ for all i,t, where α is a constant $(0 < \alpha < 1)$, we have $\|A_t\| \leq 1 - \alpha$ for all t. Therefore $\|\prod_{t=1}^{N} A_t\| \leq (1-\alpha)^N$.)

We want to determine relations between $\underline{A}$, $\bar{A}$ and G_t which can be used to achieve bounds on wildlife population in the total area in a given time period or in the long run.

Our first lemma will give the bounds on $\prod_{t=1}^{N} A_t$ in terms of the norms of $\bar{A}$ and $\underline{A}$.

<u>Lemma IV.1</u>: $\|\underline{A}^N\| \leq \|\prod_{t=1}^{N} A_t\| \leq \|\bar{A}^N\|$.

Proof: Since $0 < \underline{A} \leq A_t \leq \bar{A}$, there exists a non-negative matrix A_t' such that

$$A_t + A_t' = \bar{A},$$

hence $\prod_{t=1}^{N} A_t \leq \bar{A}^N$, thus $\|\prod_{t=1}^{N} A_t\| \leq \|\bar{A}^N\|$. Similarly $\|\prod_{t=1}^{N} A_t\| \geq \|\underline{A}^N\|$. Therefore $\|\underline{A}^N\| \leq \|\prod_{t=1}^{N} A_t\| \leq \|\bar{A}^N\|$.

For the next lemma let $\bar{\lambda}$ and $\underline{\lambda}$ be the dominant eigenvalues of $\bar{A}$ and $\underline{A}$ respectively and

$$M(\bar{A}) = \max_{i,j} \bar{a}(i,j), \quad M(\underline{A}) = \max_{i,j} \underline{a}(i,j),$$

$$m(\bar{A}) = \min_{i,j} \bar{a}(i,j), \quad m(\underline{A}) = \min_{i,j} \underline{a}(i,j).$$

<u>Lemma IV.2</u>: $\underline{\lambda}^N \frac{m(\underline{A})}{M(\underline{A})} \leq \|\underline{A}^N\| \leq \|\prod_{t=1}^{N} A_t\| \leq \|\bar{A}^N\| \leq \bar{\lambda}^N \frac{M(\bar{A})}{m(\bar{A})}$.

Proof: Since $\bar{A} > 0$, we can apply the canonical form theory in Gantmacher (1959) to get $\bar{A} = \bar{\lambda}\, \bar{Z}\, \bar{P}\, \bar{Z}^{-1}$ for some stochastic matrix $\bar{P}$. Here $\bar{\lambda} > 0$ and $\bar{\lambda}$ is the dominant eigenvalue of $\bar{A}$ and $\bar{Z} = \mathrm{diag}(\bar{z}_1, \bar{z}_2, \cdots, \bar{z}_{n-1})$, where $(\bar{z}_1, \bar{z}_2, \cdots, \bar{z}_{n-1})$ is a positive eigenvector associated with $\bar{\lambda}$. Similarly we have $\underline{A} = \underline{\lambda}\, \underline{Z}\, \underline{P}\, \underline{Z}^{-1}$. Therefore

$$\bar{A}^N = \bar{\lambda}^N \bar{Z}\, \bar{P}^N \bar{Z}^{-1},$$

$$\underline{A}^N = \underline{\lambda}^N \underline{Z}\, \underline{P}^N \underline{Z}^{-1}.$$

Now, by application of known bounds on the components of $\underline{Z}$ and $\bar{Z}$, Ostrowski (1960), we have

$$\|\bar{A}^N\| = \bar{\lambda}^N \|\bar{Z}\ \bar{P}^N \bar{Z}^{-1}\| \leq \bar{\lambda}^N \frac{M(\bar{A})}{m(\bar{A})},$$

and
$$\|\underline{A}^N\| \geq \underline{\lambda}^N \frac{m(\underline{A})}{M(\underline{A})}.$$

By Lemma IV.1 we thus have

$$\underline{\lambda}^N \frac{m(\underline{A})}{M(\underline{A})} \leq \|\underline{A}^N\| \leq \|\prod_{t=1}^{N} A_t\| \leq \|\bar{A}^N\| \leq \bar{\lambda}^N \frac{M(\bar{A})}{m(\bar{A})}. \tag{1}$$

Remark: It is known that $\bar{\lambda} \geq \underline{\lambda}$, Gantmacher (1959).

From the fact that

$$P_t G_t = \begin{pmatrix} 1 & 0 \\ Q_t & r_t A_t \end{pmatrix},$$

we have

$$\prod_{t=1}^{N} P_t G_t = \prod_{t=1}^{N} \begin{pmatrix} 1 & 0 \\ Q_t & r_t A_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ Q^*(N) & \prod_{t=1}^{N} r_t A_t \end{pmatrix},$$

hence we concentrate on $\prod_{t=1}^{N} r_t A_t$. We ask: How can we avoid a zero population or an infinite population in the long run?

<u>Theorem IV.3</u>: If $r_t \geq \frac{1}{\underline{\lambda}}$ for all t, then

$$\inf_N \Big(\prod_{t=1}^{N} r_t\Big) \Big\| \prod_{t=1}^{N} A_t \Big\| > 0.$$

If $r_t \leq \frac{1}{\bar{\lambda}}$ for all t, then

$$\sup_N \Big(\prod_{t=1}^{N} r_t\Big) \Big\| \prod_{t=1}^{N} A_t \Big\| < \infty.$$

Proof: By (1) we have

$$\Big(\prod_{t=1}^{N} r_t\Big) \underline{\lambda}^N \frac{m(\underline{A})}{M(\underline{A})} \leq \prod_{t=1}^{N} r_t \Big\| \prod_{t=1}^{N} A_t \Big\| \leq \Big(\prod_{t=1}^{N} r_t\Big) \bar{\lambda}^N \frac{M(\bar{A})}{m(\bar{A})} .$$

If $r_t \geq \frac{1}{\underline{\lambda}}$ for all t, then

$$0 < \frac{m(\underline{A})}{M(\underline{A})} \le \left(\prod_{t=1}^{N} r_t\right) \left\| \prod_{t=1}^{N} A_t \right\|, \quad \text{that is,}$$

$$\inf_N \left(\prod_{t=1}^{N} r_t\right) \left\| \prod_{t=1}^{N} A_t \right\| > 0.$$

Similarly, if $r_t \le \frac{1}{\bar{\lambda}}$ for all t, then

$$\sup_N \left(\prod_{t=1}^{N} r_t\right) \left\| \prod_{t=1}^{N} A_t \right\| < \infty.$$

Remark: If there exists T such that $A_t = A_{t+1}$ for all $t \geq T$, then it can be shown that neither the zero population nor the infinite population will happen. The reason is that since $A_t = A_{t+1}$ for all $t \geq T$, then WLOG, we may assume $\underline{A} = A_t = \bar{A}$ for all t. Then $\underline{\lambda} = \bar{\lambda}$, and using $\underline{\lambda} = \bar{\lambda} = r_t$ for all t, it is easy to see that neither the zero population nor the infinite population will happen in the long run.

Next suppose we know an initial distribution given by $\zeta_0 = [\zeta_1, \zeta_2, \cdots, \zeta_{n-1}]$. What can we say about the expected population after N seasons or in the long run by using $r_t = \frac{1}{\bar{\lambda}}$ for all t or $r_t = \frac{1}{\underline{\lambda}}$ for all t.

<u>Theorem IV.4</u>: If $r_t = \frac{1}{\bar{\lambda}}$ for all t, then

$$\underline{\zeta_0}(n-1)\,\frac{m(\underline{A})}{M(\underline{A})}\left(\frac{\underline{\lambda}}{\bar{\lambda}}\right)^N \leq \left\|\zeta_0\left(\prod_{t=1}^{N} r_t A_t\right)\right\| \leq \|\zeta_0\|\frac{M(\bar{A})}{m(\bar{A})},$$

where $\underline{\zeta_0} = \min\{\zeta_1, \zeta_2, \cdots, \zeta_{n-1}\}$.

If $r_t = \frac{1}{\underline{\lambda}}$ for all t, then

$$\underline{\zeta_0}(n-1)\,\frac{m(\underline{A})}{M(\underline{A})} \leq \left\|\zeta_0\left(\prod_{t=1}^{N} r_t A_t\right)\right\| \leq \|\zeta_0\|\frac{M(\bar{A})}{m(\bar{A})}\left(\frac{\bar{\lambda}}{\underline{\lambda}}\right)^N,$$

where $\underline{\zeta_0} = \min\{\zeta_1, \zeta_2, \cdots, \zeta_{n-1}\}$.

Proof: $\left(\prod_{t=1}^{N} r_t\right)\underline{A}^N \leq \prod_{t=1}^{N} r_t A_t \leq \left(\prod_{t=1}^{N} r_t\right)\bar{A}^N$, if $r_t = \frac{1}{\bar{\lambda}}$ for all t then

$$\underline{\zeta_0}(n-1)\frac{m(\underline{A})}{M(\underline{A})}\left(\frac{\underline{\lambda}}{\bar{\lambda}}\right)^N \leq \|\zeta_0\underline{P}^N\|\frac{m(\underline{A})}{M(\underline{A})}\left(\frac{\underline{\lambda}}{\bar{\lambda}}\right)^N$$

$$\leq \|\zeta_0\Big(\prod_{t=1}^{N} r_t\Big)\underline{A}^N\| \quad (\text{for} \quad \underline{A}^N = \underline{\lambda}^N\underline{Z}\,\underline{P}^N\underline{Z}^{-1})$$

$$\leq \|\zeta_0\Big(\prod_{t=1}^{N} r_t A_t\Big)\| \leq \|\zeta_0\Big(\prod_{t=1}^{N} r_t\Big)\bar{A}^N\| \leq \|\zeta_0\|\frac{M(\bar{A})}{m(\bar{A})},$$

i.e. $\underline{\zeta_0}(n-1)\frac{m(\underline{A})}{M(\underline{A})}\left(\frac{\underline{\lambda}}{\bar{\lambda}}\right)^N \leq \|\zeta_0\Big(\prod_{t=1}^{N} r_t A_t\Big)\| \leq \|\zeta_0\|\frac{M(\bar{A})}{m(\bar{A})}$.

Similarly, if $r_t = \frac{1}{\underline{\lambda}}$ for all t, then

$$\underline{\zeta_0}(n-1)\frac{m(\underline{A})}{M(\underline{A})} \leq \|\zeta_0\Big(\prod_{t=1}^{N} r_t A_t\Big)\| \leq \|\zeta_0\|\frac{M(\bar{A})}{m(\bar{A})}\left(\frac{\bar{\lambda}}{\underline{\lambda}}\right)^N.$$

<u>Remark</u>:

i) In Theorem IV.4 if $N \to \infty$ then we can prevent the population from going to infinity by using $r_t = \frac{1}{\bar{\lambda}}$ for all t and thus

$$\left\|\zeta_0\left(\prod_{t=1}^{N} r_t A_t\right)\right\| \leq \|\zeta_0\|\frac{M(\bar{A})}{m(\bar{A})} \; .$$

And we also can prevent the population from going to zero by using $r_t = \frac{1}{\underline{\lambda}}$ for all t, and thus

$$\left\|\zeta_0\left(\prod_{t=1}^{N} r_t A_t\right)\right\| \geq \underline{\zeta_0}(n-1)\frac{m(\underline{A})}{M(\underline{A})} \; .$$

ii) If $\bar{\lambda}$ and $\underline{\lambda}$ are close, then we will get a better estimate of

$$\left\|\zeta_0\left(\prod_{t=1}^{N} r_t A_t\right)\right\| .$$

iii) $$\max\left\{0, \min_i \sum_j \bar{a}(i,j) - \|\underline{A}\|\right\} \leq \bar{\lambda} - \underline{\lambda}$$

$$\leq \|\bar{A}\| - \min_i \sum_j \underline{a}(i,j) .$$

The following theorem displays some easily computed bounds for

$$\zeta_0\left(\prod_{t=1}^{N} r_t\right)\bar{A}^N \quad \text{and} \quad \zeta_0\left(\prod_{t=1}^{N} r_t\right)\underline{A}^N .$$

<u>Theorem IV.5</u>: If $\prod_{t=1}^{N} r_t = (\bar{\lambda}^N)^{-1}$, then

$$\|\zeta_0\bar{P}^N\|\frac{m(\bar{A})}{M(\bar{A})} \leq \left\|\zeta_0\left(\prod_{t=1}^{N} r_t\right)\bar{A}^N\right\| \leq \|\zeta_0\|\frac{M(\bar{A})}{m(\bar{A})} ,$$

and if $\prod_{t=1}^{N} r_t = (\underline{\lambda}^N)^{-1}$, then

$$\|\zeta_0\underline{P}^N\|\frac{m(\underline{A})}{M(\underline{A})} \leq \left\|\zeta_0\left(\prod_{t=1}^{N} r_t\right)\underline{A}^N\right\| \leq \|\zeta_0\|\frac{M(\underline{A})}{m(\underline{A})} .$$

Proof: It is straightforward by first noting that $\bar{A}^N = \bar{\lambda}^N \bar{Z}\, \bar{P}^N \bar{Z}^{-1}$ and $\underline{A}^N = \underline{\lambda}^N \underline{Z}\, \underline{P}^N \underline{Z}^{-1}$, and using Lemma IV.2.

<u>Example IV.1</u>: Let

$$A_t = \begin{pmatrix} \frac{1}{2} + \frac{1}{20t}, & \frac{1}{3} - \frac{1}{20t} \\ \\ \frac{1}{3} + \frac{1}{10t}, & \frac{1}{4} - \frac{1}{10t} \end{pmatrix}, \quad t = 1,2,3,\cdots \quad .$$

Then

$$\begin{pmatrix} \frac{1}{2} & \frac{17}{60} \\ \\ \frac{1}{3} & \frac{3}{20} \end{pmatrix} = \underline{A} \leq A_t \leq \bar{A} = \begin{pmatrix} \frac{11}{20} & \frac{1}{3} \\ \\ \frac{13}{30} & \frac{1}{4} \end{pmatrix} .$$

The dominant eigenvalue $\bar{\lambda}$ of $\bar{A}$ is

$$\bar{\lambda} = \frac{288 + \sqrt{82944}}{720} \doteq \frac{4}{5} ,$$

and the dominant eigenvalue $\underline{\lambda}$ of $\underline{A}$ is

$$\underline{\lambda} = \frac{117 + \sqrt{16209}}{360} \doteq \frac{61}{90} .$$

Therefore $\bar{\lambda} \doteq \frac{4}{5} > \frac{61}{90} \doteq \underline{\lambda}$. Now $m(\bar{A}) = \frac{1}{4}$, $M(\bar{A}) = \frac{11}{20}$, $m(\underline{A}) = \frac{3}{20}$, $M(\underline{A}) = \frac{1}{2}$, thus

$$\underline{\lambda}^N \frac{m(\underline{A})}{M(\underline{A})} \doteq \left(\frac{61}{90}\right)^N \frac{\frac{3}{20}}{\frac{1}{2}} = \left(\frac{61}{90}\right)^N \left(\frac{3}{10}\right),$$

$$\bar{\lambda}^N \frac{M(\bar{A})}{m(\bar{A})} \doteq \left(\frac{4}{5}\right)^N \frac{\frac{11}{20}}{\frac{1}{4}} = \left(\frac{4}{5}\right)^N \left(\frac{11}{5}\right) .$$

This implies

$$\left(\frac{61}{90}\right)^N \left(\frac{3}{10}\right) \leq \|\underline{A}^N\| \leq \|\prod_{t=1}^{N} A_t\| \leq \|\bar{A}^N\| \leq \left(\frac{4}{5}\right)^N \left(\frac{11}{5}\right).$$

If the growth factor $r_t = \underline{\lambda}^{-1} \doteq \frac{90}{16}$ for all t, then

$$\left(\prod_{t=1}^{N} r_t\right)\left\|\prod_{t=1}^{N} A_t\right\| \geq \frac{3}{10} \neq 0.$$

If the growth factor $r_t = \bar{\lambda}^{-1} \doteq \frac{5}{4}$ for all t, then

$$\left(\prod_{t=1}^{N} r_t\right)\left\|\prod_{t=1}^{N} A_t\right\| \leq \frac{11}{5} \neq \infty.$$

Next assume that we have initial population vector $\zeta_0 = (2000,3000)$, and suppose we are going to predict the population in the long run.

i) If $r_t = \frac{1}{\bar{\lambda}} \doteq \frac{5}{4}$ for all t, we get

$$\left\|\zeta_0 \prod_{t=1}^{N} (r_t A_t)\right\| \leq (5000) \frac{\frac{11}{20}}{\frac{1}{4}} = 11000 \quad \text{for} \quad N = 1,2,3,\cdots .$$

Thus, if we bound the population away from infinity, then the population will reach at most 11000 when the original population is 5000.

ii) If $r_t = \frac{1}{\underline{\lambda}} \doteq \frac{90}{61}$ for all t, we get

$$\underline{\zeta_0}(n-1)\frac{m(\underline{A})}{M(\underline{A})} \leq \|\zeta_0 \prod_{t=1}^{N} (r_t A_t)\| \leq \|\zeta_0\|\frac{M(\bar{A})}{m(\bar{A})}\left(\frac{\bar{\lambda}}{\underline{\lambda}}\right)^N,$$

that is,

$$1200 \leq \|\zeta_0 \prod_{t=1}^{N} (r_t A_t)\| \leq 5000\left(\frac{11}{20}\right)\left(\frac{4}{1}\right)\left(\frac{72}{61}\right)^N .$$

Thus

$$1200 \leq \|\zeta_0 \prod_{t=1}^{N} (r_t A_t)\| \leq 11000\left(\frac{72}{61}\right)^N .$$

These are admittedly crude bounds, but this is a consequence of the small amount of information used.

C. Wildlife Model: Use of $\underline{A}$, $\bar{A}$ with Diagonal Normalization

In place of

$$G_t = \begin{pmatrix} 1 & 0 \\ 0 & r_t I_{n-1} \end{pmatrix},$$

let us allow different growth factors, say

$$G_t = \begin{pmatrix} 1 & & & & 0 \\ & r_t(1) & & & \\ & & r_t(2) & & \\ & & & \ddots & \\ 0 & & & & r_t(n-1) \end{pmatrix} \quad \text{and}$$

$$H_t = \begin{pmatrix} r_t(1) & & & 0 \\ & r_t(2) & & \\ & & \ddots & \\ 0 & & & r_t(n-1) \end{pmatrix}.$$

We now give a reduction to the scalar case which, although it sacrifices more information, does yield bounds. Let A_t, $\underline{A}$, $\bar{A}$, $\underline{\lambda}$, $\bar{\lambda}$, H_t, $r_t(i)$ be as in section B, so in particular, $0 \leq \underline{\lambda} < \bar{\lambda} < 1$ and $r_t(i) \geq 1$. Now set $\bar{r}_t = \max_i r_t(i)$. Then

$$A_t H_t = \bar{r}_t A_t \begin{pmatrix} \frac{r_t(1)}{\bar{r}_t} & & & 0 \\ & \frac{r_t(2)}{\bar{r}_t} & & \\ & & \ddots & \\ 0 & & & \frac{r_t(n-1)}{\bar{r}_t} \end{pmatrix}$$

$$= \bar{r}_t B_t$$

where

$$b_t(\sigma,\tau) = a_t(\sigma,\tau)\left[\frac{r_t(\tau)}{\bar{r}_t}\right]$$

$$= a_t(\sigma,\tau)\rho_t(\tau) \quad \text{where} \quad \rho_t(\tau) = \frac{r_t(\tau)}{\bar{r}_t} .$$

Now $B_t \leq A_t$ implies $\bar{B} \leq \bar{A}$, $\underline{B} \leq \underline{A}$. Clearly $1 = \sup\limits_{\tau,t} \rho_t(\tau)$, let $\underline{\rho} = \inf\limits_{\tau,t} \rho_t(\tau)$. Then $\bar{A} \geq \bar{B} \geq \underline{\rho}\,\bar{A}$, therefore $\bar{\lambda} \geq \bar{\mu} \geq \underline{\rho}\,\bar{\lambda}$ where $\bar{\mu}$ is the dominant eigenvalue of $\bar{B}$. Similarly $\underline{\lambda} \geq \underline{\mu} \geq \underline{\rho}\,\underline{\lambda}$ where $\underline{\mu}$ is the dominant eigenvalue of $\underline{B}$, since $\underline{A} \geq \underline{B} \geq \underline{\rho}\,\underline{A}$.

Now $\prod\limits_{t=1}^{N} (A_t G_t) = \prod\limits_{t=1}^{N} (\bar{r}_t B_t)$ implies

$$\underline{\lambda}^N \left(\prod_{t=1}^{N} \bar{r}_t\right) \underline{\rho}^N \frac{m(\underline{A})}{M(\underline{A})} \leq \left(\prod_{t=1}^{N} \bar{r}_t\right) \underline{\rho}^N \|\underline{A}^N\| \leq \left\| \prod_{t=1}^{N} (\bar{r}_t \underline{B}) \right\|$$

$$\leq \left\| \prod_{t=1}^{N} \bar{r}_t B_t \right\|$$

$$\leq \left\| \prod_{t=1}^{N} \bar{r}_t \bar{B} \right\|$$

$$\leq \left(\prod_{t=1}^{N} \bar{r}_t\right) \cdot \|\bar{A}^N\|$$

$$\leq \left(\prod_{t=1}^{N} \bar{r}_t\right) \cdot \frac{M(\bar{A})}{m(\bar{A})} \bar{\lambda}^N .$$

From the above discussion we have the following theorem:

<u>Theorem IV.6</u>:

i) If $\prod_{t=1}^{N} \bar{r}_t \leq \frac{1}{\bar{\lambda}^N}$ then $\left\| \prod_{t=1}^{N} (A_t G_t) \right\| \leq \frac{M(\bar{A})}{m(\bar{A})}$

ii) If $\prod_{t=1}^{N} \bar{r}_t \geq \frac{1}{(\underline{\rho\lambda})^N}$ then $\left\| \prod_{t=1}^{N} (A_t G_t) \right\| \geq \frac{m(\underline{A})}{M(\underline{A})}$

D. Weakly Ergodic Theory for Diagonal Normalization

Our discussion of normalization suggests that the ergodic theory in Isaacson and Madsen (1974) can be pushed through for diagonal normalization. Here we restrict ourselves to the finite state case in describing the results. We are indebted to the above paper for the theory and methods which we adapt to the case of a normalized sequence of matrices.

<u>Definition IV.1</u>: Let $\{A_t\}$ be a sequence of non-negative matrices and let M be the family of all starting probability vectors. The sequence $\{A_t\}$ will be called normalized weakly ergodic if for each pair $\zeta_0, \eta_0 \in M$ there exist sequences of positive constants $K(\zeta_0, m, n)$ and $K(\eta_0, m, n)$ such that for all m

$$\sup_{\zeta_0, \eta_0 \in M} \sum_y |f^*_{m,n}(y) - g^*_{m,n}(y)| \to 0 \quad \text{as} \quad n \to \infty$$

and

$$\sum_y f^*_{m,n}(y) \not\to 0,$$

where

$$f^*_{m,n}(y) = \frac{\left(\zeta_0 \prod_{t=m}^{n} A_t G_t\right)(y)}{K(\zeta_0, m, n)} \quad ,$$

$$g^*_{m,n}(y) = \frac{\left(\eta_0 \prod_{t=m}^{n} A_t G_t\right)(y)}{K(\eta_0, m, n)},$$

and G_t is defined as in section C.

The following lemmas will be useful in the proof of the main theorem of this section.

<u>Lemma IV.7</u>: If $\sum_x g(x)$ exists and $h(x)$ satisfies $0 \le 1 - \epsilon \le h(x) \le 1 + \epsilon$. Then

$$\left|\sum_x g(x)h(x)\right| \le \left|\sum_x g(x)\right| + \epsilon \sum |g(x)|$$

Proof: If we write $g(x) = g^+(x) - g^-(x)$ and $|g(x)| = g^+(x) + g^-(x)$ and use the bounds on $h(x)$, the proof is easy.

Let $z_t = [z_t(1), z_t(2), \cdots]$ be a positive right eigenvector of A_t (for A_t see section B) corresponding to the dominant eigenvalue $\lambda_t > 0$, then

(1) $A_t = \lambda_t Z_t P_t Z_t^{-1}$, where $Z_t = \mathrm{diag}(z_t(1), z_t(2), \cdots)$

and P_t is a stochastic matrix. For further reference we state the following modifications of properties assumed in Isaacson and Madsen (1974).

Property I: There exist right eigenvectors Z_t corresponding to $\lambda_t > 0$ such that

i) $0 < b \leq z_t(i) \leq B < \infty$,

ii) $r_t = \sup_i \left| \frac{z_t(i)}{z_{t-1}(i)} \cdot \frac{r_{t-1}(i)}{\bar{r}_{t-1}} - 1 \right|$ satisfies

$$\sum_{t=2}^{\infty} r_t < \infty, \quad \text{where} \quad \bar{r}_t = \max_i r_t(i).$$

Property II: The sequence of stochastic matrices $\{P_t\}$ defined by (1) is weakly ergodic.

Lemma IV.8: Let $\{A_t\}$ be a sequence of non-negative matrices satisfying Properties I and II. Then, given $\epsilon > 0$ and starting distributions ζ_0 and η_0, there are sequences of normalizing constants $\{d_n(\epsilon)\}$ and

$\{e_n(\epsilon)\}$ such that for $n \geq N(\epsilon)$

$$\sum_y |f^*_{1,n}(y) - g^*_{1,n}(y)| < \epsilon.$$

Proof: Define $f_n(y) = f_{1,n}(y)$. Since $\sum_{n=2}^{\infty} r_n < \infty$, it follows that $\prod_{n=2}^{\infty} (1 + r_n)$ converges. Hence, given ν such that $0 < \nu < 1$, there exists $N_1 = N_1(\nu)$ such that

$$\prod_{n=N_1+1}^{M} (1 + r_n) \leq \prod_{N=N_1+1}^{\infty} (1 + r_n) < 1 + \nu$$

for all $M > N_1 + 1$. Furthermore

$$\prod_{n=N_1+1}^{M} (1 - r_n) \geq \prod_{n=N_1+1}^{\infty} (1 - r_n) > 1 - \nu.$$

Also, since $\{P_t\}$ is weakly ergodic, given m and $\nu > 0$, there exists an $N_2 = N_2(\nu, m)$ such that for $n \geq N_2$,

$$\delta\left(\prod_{t=m}^{n} P_t\right) < \nu .$$

Let $\nu = \frac{\epsilon b}{4}$, $m = N_1(\nu)$ and $N(\epsilon) = N_2(\nu, m)$. If ζ_0 and η_0 are given starting distributions, define $\{d_n\}$ as follows:

$$d_n = \begin{cases} 1 & \text{if } n < m \\ \prod_{i=m}^{n} \lambda_t \bar{r}_t \sum_y \left(\zeta_0\left(\prod_{t=1}^{m-1} A_t G_t\right)\right)(y) z_m(y) & \text{if } n \geq m. \end{cases}$$

then $d_n > 0$, for all n. (For Property I.i).)

$$e_n = \begin{cases} 1 & \text{if } n < m \\ \prod_{i=m}^{n} \lambda_t \bar{r}_t \sum_y \left(\eta_0\left(\prod_{t=1}^{m-1} A_t G_t\right)\right)(y) z_m(y) & \text{if } n \geq m. \end{cases}$$

then $e_n > 0$ for all n. (For Property I.i).) Now for

$n \geq N(\epsilon) = N_2(\nu,m)$, since $n > m$, we can write

$$\left(\zeta_0\left(\prod_{t=1}^{n} A_t G_t\right)\right)(y) = \sum_{i_n} \cdots \sum_{i_m} \left(\zeta_0\left(\left(\prod_{t=1}^{m-1} A_t G_t\right)(i_m)\right)\right)((A_m G_m)$$

$$(i_m, i_{m+1}))\cdots(A_n G_n(i_n, y))$$

$$= \sum_{i_n} \cdots \sum_{i_m} \left(\zeta_0\left(\left(\prod_{t=1}^{m-1} A_t G_t\right)(i_m)\right)\right)$$

$$[\lambda_m Z_m P_m Z_m^{-1} G_m(i_m, i_{m+1}) \cdots$$

$$\lambda_n Z_n P_n Z_n^{-1} G_n(i_n, y)]$$

$$= \sum_{i_n} \cdots \sum_{i_m} \left(\zeta_0\left(\left(\prod_{t=1}^{m-1} A_t G_t\right)(i_m)\right)\right)\left(\prod_{t=m}^{n} \lambda_t\right)$$

$$z_m(i_m) P_m(i_m, i_{m+1}) z_m^{-1}(i_{m+1})) G_m(i_{m+1}) \cdots$$

$$(z_n(i_n) P_n(i_n, y) z_n^{-1}(y)) G_n(y)$$

$$= \sum_{i_n} \cdots \sum_{i_m} \left(\zeta_0 \left(\left(\prod_{t=1}^{m-1} A_t G_t \right) (i_m) \right) \right)$$

$$\left(\prod_{t=m}^{n} \lambda_t \bar{r}_t (z_m (i_m)) \right) P_m (i_m, i_{m+1}) \cdots$$

$$P_n (i_m, y) \left(\prod_{t=m+1}^{n} \frac{z_t (i_t)}{z_{t-1} (i_t)} \cdot \frac{r_{t-1} (i_t)}{\bar{r}_{t-1}} \right)$$

$$z_n^{-1} (y) B_n (y),$$

(for B_n see section C) and define

$$f_{n-1}^{**} (y) = \left(\zeta_0 \left(\prod_{t=1}^{m-1} A_t G_t \right) \right) (y) z_m (y) / \sum_{y} \left(\zeta_0 \prod_{t=1}^{m-1} A_t G_t \right) (y) z_m (y) .$$

Then

$$f_n^*(y) = \frac{\left(\zeta_0 \prod_{t=1}^{n} A_t G_t\right)(y)}{d_n}$$

$$= \sum_{i_n} \cdots \sum_{i_m} f_{m-1}^{**}(i_m) P_m(i_m, i_{m+1}) \cdots P_n(i_n, y)$$

$$\left\{ \prod_{t=m+1}^{n} \frac{z_t(i_t)}{z_{t-1}(i_t)} \cdot \frac{r_{t-1}(i_t)}{\bar{r}_{t-1}} \right\} z_n^{-1}(y) B_n(y) .$$

A similar expression can be given for $g_n^*(y)$. Hence

$$\left\| \left(\zeta_0\left(\prod_{t=1}^{n} A_t G_t\right)\right)(y)/d_n - \left(\eta_0\left(\prod_{t=1}^{n} A_t G_t\right)\right)(y)/e_n \right\|$$

$$= \sum_{y} \left| \sum_{i_n} \cdots \sum_{i_m} (f_{m-1}^{**}(i_m) - g_{m-1}^{**}(i_m)) P_m(i_m, i_{m+1}) \cdots P_n(i_n, y) \right.$$

$$\left. \left\{ \prod_{t=m+1}^{n} \frac{z_t(i_t)}{z_{t-1}(i_t)} \cdot \frac{r_{t-1}(i_t)}{\bar{r}_{t-1}} \right\} \right| z_n^{-1}(y) B_n(y) .$$

Since $1 - \nu \leq \prod_{t=m+1}^{n} \frac{z_t(i_t)}{z_{t-1}(i_t)} \cdot \frac{r_{t-1}(i_t)}{\bar{r}_{t-1}} \leq 1 + \nu$, by

Lemma V.7 we have

$$\left\| \left(\zeta_0\left(\prod_{t=1}^{n} A_t G_t\right)(y)/d_n - \left(\eta_0\left(\prod_{t=1}^{n} A_t G_t\right)(y)/e_n\right\|\right.\right.$$

$$\leq \sum_y \left|\sum_{i_m} [f^{**}_{m-1}(i_m) - g^{**}_{m-1}(i_m)] P_{m,n}(i_m,y)\right| z_n^{-1}(y) B_n(y)$$

$$+ \nu \sum_y \sum_{i_m} \left|f^{**}_{m-1}(i_m) - g^{**}_{m-1}(i_m)\right| P_{m,n}(i_m,y) z_n^{-1}(y) B_n(y) \cdots (2)$$

The first term of (2) becomes

$$\sum_y \left|\sum_{i_m} (f^{**}_{m-1}(i_m) - g^{**}_{m-1}(i_m)) P_{m,n}(i_m, y)\right| z_n^{-1}(y) B_n(y)$$

$$\leq \frac{1}{b} \left|\sum_{i_m} f^{**}_{m-1}(i_m) P_{m,n}(i_m,y) - g^{**}_{m-1}(i_m) P_{m,n}(i_m,y)\right|$$

(for $B_n(y) \leq 1$)

$$\leq \frac{1}{b} \left|\max_{i_m} P_{m,n}(i_m,y) - \min_{i_m} P_{m,n}(i_m,y)\right|$$

$$\leq \frac{2}{b}\,\delta(P_{m,n}) \leq \frac{2\nu}{b} .$$

Now consider the second term of (2). It is less than or equal to

$$\frac{\nu}{b} \sum_{y} \sum_{i_m} \left|f^{**}_{m-1}(i_m) - g^{**}_{m-1}(i_m)\right| P_{m,n}(i_m,y) \leq \frac{2\nu}{b} .$$

Thus $$\sum_{y} \left|f^{*}_{n}(y) - g^{*}_{n}(y)\right| \leq \frac{2\nu}{b} + \frac{2\nu}{b} = \frac{4\nu}{b} = \epsilon .$$

Theorem IV.9: If $\{A_t\}$ is a sequence of non-negative square matrices satisfying I and II, then $\{A_t\}$ is normalized weakly ergodic.

Proof: Let ζ_0 and η_0 be any starting distributions. It is sufficient to consider the case $m = 1$, since for any other m the arguments are identical.

Let $\{\epsilon_i\}$ be a sequence of constants decreasing to zero. By Lemma V.8 for each i there exist sequences of constants $\{d_n(\epsilon_i)\}$ and $\{e_n(\epsilon_i)\}$ such that $n \geq N(\epsilon_i)$ implies

$$\sum_y |f_n^*(y) - g_n^*(y)| < \epsilon_i .$$

Without loss of generality, assume that $\{N(\epsilon_i)\}$ forms an increasing sequence. Since $\{N(\epsilon_i)\}$ does not depend on ζ_0 or η_0, define

$$K(\zeta_0,1,n) = \begin{array}{ll} d_n(\epsilon_1) & n \leq N(\epsilon_2) \\ & \\ d_n(\epsilon_i) & N(\epsilon_i) < n \leq N(\epsilon_{i+1}) \end{array} ,$$

and define $K(\eta_0,1,n)$ similarly using $\{e(\epsilon_i)\}$. These sequences of constants can be used to show that $\{A_t\}$ is normalized weakly ergodic directly from the definition. If $\epsilon > 0$ is given, there is some i such that $\epsilon_i < \epsilon$ and for any $n > N(\epsilon_i)$,

$$\sum_y |f_n^*(y) - g_n^*(y)| < \epsilon_i < \epsilon$$

independently of the choice of ζ_0 and η_0.

It remains to show that

$$\sum_y f_n^*(y) \nrightarrow 0.$$

For $m = N_1(\nu)$,

$$\sum_y f_n^*(y) = \sum_y \sum_{i_n} \cdots \sum_{i_m} f_{m-1}^{**}(i_m) P_m(i_m, i_{m+1}) \cdots P_n(i_n, y)$$

$$\left\{ \prod_{t=m+1}^{n} \frac{z_t(i_t)}{z_{t-1}(i_t)} \frac{r_{t-1}(i_t)}{\bar{r}_{t-1}} \right\} z_n^{-1}(y) B_n(y)$$

$$\geq \frac{1}{Bk} \prod_{t=m+1}^{n} (1 - r_j) \Big(\sum_y \sum_{i_n} \cdots \sum_{i_m} f_{m-1}^{**}(i_m) P_m(i_m, i_{m+1})$$

$$\cdots P_n(i_n, y) \Big) \text{ (for } 0 < k \leq B_n(y) = 1)$$

$$= \frac{1}{Bk} \prod_{j=m+1}^{n} (1 - r_j) \geq \frac{1 - \nu}{B} > 0$$

for all n.

Note: Most of the proofs in this section are conducted as in Isaacson and Madsen (1974). In the latter paper the results apply directly to scalar normalization case, but if we change the condition in Property I.ii) from Isaacson's and Madsen's paper to our present Property I.ii) then all works for diagonal normalization.

It may be noted that the sequence $\{d_n\}$ as defined in Isaacson and Madsen (1974), that is without the $\bar{r}_t$'s, corresponds to a scalar normalization by $w_n = d_n/d_{n-1}$. Thus our modified $\{d_n\}$ is a device for handling a perturbation of this scalar normalization.

E. An Increasing Growth Stock Model in Forestry

The commercial timber resource in a forest may be measured in terms of two types of growth stock: regeneration and intermediate.

Growth stock is the inventory of wood fiber in the forest and is universally measured in cubic feet of volume. It includes trees below sawlog size, in recognition of the diverse uses (other than lumber) of wood fiber.

Regeneration stock is measured in areas where mature and overmature trees are present in sufficient volume for commercial harvest. In such areas the overstory, when harvested, will release younger stock, and although this stock was counted as part of the regeneration stock it will change its classification (after the harvest) to intermediate stock as defined next.

Intermediate stock is measured in areas where only immature trees occur. Its harvest is by thinning and other timber improvement practices. Note again that a harvest in the regeneration stock releases some growing stock to the intermediate category. On the other hand, much of the intermediate stock will enter the regeneration category as a result of growth during a transition period (see definition of cutting period below).

If a forest is not now providing a sustained yield, that is, it is not yielding a periodic harvest equal to its periodic growth maximized subject to environmental constraints and optimal silvicultural practices, then it is desirable to manage the forest by a long range plan whose purpose is to achieve sustained yield. There exist standard procedures applied by professional foresters. But to illustrate some of our "growth factor" theory we consider a different and very idealized approach which emphasizes a sustained growth stock.

Our goal is to stabilize the total growth stock. We assume the forest is an uneven-age type on which selective cutting (in contrast to clearcutting) is the principal method of harvest. We also assume that the present condition of the forest is one of inadequate growth stock and excess of mature and overmature trees, and capable of receiving management practice which if periodically applied will achieve a monotonic increase of growth stock to a silviculturally and ecologically desirable level.

For transition period we use the "cutting" period (say 20 years), that is the time span during which a constant harvest and timber improvement management plan applies. We shall let t be measured in units of 20 years and proceed in the spirit of the following quote from the Timber Management Plan of the Dixie National Forest (1976, p.28).

"There is a need to convert mature and overmature timber to younger, thriftier stands. At the same time, there is an opportunity to treat the immature stands to gain the most from their potential. Priority will be given to the release of younger understory stands through removal of mature and overmature stands. Maximum effort will be directed toward accomplishing thinning through commercial disposal. Sanitation cutting in sawtimber stands may be accomplished by removing high risk trees in stands adjacent to the harvest cut areas."

The heart of our approach lies in the transient block $A_t = [P_t(i,j)]$; $i,j = 2,3$ of the 3×3 stochastic matrix

$$P_t = \left(\begin{array}{c:cc} 1 & 0 & 0 \\ \hdashline p_t(2,1) & & \\ p_t(3,1) & & A_t \end{array}\right)$$

At the start of the $t^{\underline{th}}$ transition period we describe the volume of growth stock as follows:

$x_t(1)$ = cumulative volume of growth stock <u>harvested</u> <u>or</u> <u>lost</u> because of fire, insects, disease and mortality.

$x_t(2)$ = volume of growth stock in the <u>regeneration category</u>.

$x_t(3)$ = volume of growth stock in the <u>intermediate category</u>.

The underlined phrases will be our <u>states</u>, referred to as 1, 2, 3. At the end of a cutting period, these volumes redistribute themselves in accordance with

$p_t(i,j)$ = fraction of $x_t(i)$ that changes to type j during $t^{\underline{th}}$ period. We shall abbreviate this by p_{ij} when there is no confusion.

In detail,

p_{21} = fraction of regeneration volume harvested,

p_{22} = fraction of regeneration volume in unchanged state,

p_{23} = fraction of regeneration volume changed to intermediate state (due to regeneration harvest),

p_{31} = fraction of intermediate volume harvested,

p_{32} = fraction of intermediate volume that grows into regeneration state,

p_{33} = fraction of intermediate volume in unchanged state.

Set r_i = 1 + fractional increase in volume of type i due to growth during $t^{\underline{th}}$ period

= growth factor.

r_i is assumed independent of t, although it may improve with t; clearly $r_1 = 1$.

Set $\vec{x}_0 = (0, x_0(2), x_0(3))$, $\vec{y}_0 = (x_0(2), x_0(3))$,

$\vec{x}_t = (x_t(1), x_t(2), x_t(3))$, $\vec{y}_t = (x_t(2), x_t(3))$,

$G = \text{diag}(r_1, r_2, r_3)$, $H = \text{diag}(r_2, r_3)$.

Then $\vec{x}_0 \prod_{t=1}^{N} (P_t G)$ is volume distribution after N periods and $\vec{y}_0 \prod_{t=1}^{N} (A_t H)$ is volume distribution of types 2, 3 after N periods.

Set $B_t = A_t H$, $E = (1,1)^T$. Given $\vec{y}_0$ and H, it is management's purpose to select A_t so as to stabilize $\vec{y}_0 \prod_{t=1}^{N} B_t E$, subject to realistic constraints. While a regeneration harvest transforms part of the regeneration growth stock from state 2 to 3, a certain fraction of the intermediate growth stock is passing to state 3 and some harvesting (mainly thinning) and other removals (e.g. sanitary) are occurring in them every period.

If we start with the unsatisfactory growth stock situation quoted above from the DNFTMP, then an initially large regeneration harvest will gradually decrease while the intermediate harvest increases. (See acreage figures in table in Appendix. Note that this table does not give growth stock volumes.) Since p_{21} is usually maximized, we may assume

$$p_{23} > p_{22},$$

that is, the new intermediate volume released by harvest exceeds the residual mature growth stock. And to reflect replacement of the regeneration harvest from the intermediate growth stock of young adult trees, we assume

$$p_{32} > p_{33}.$$

There is more flexibility in the imposition of monotone assumptions on the p_{ij}, depending as this does on management options. Since p_{21} is $\downarrow$, we may take $p_{22}\uparrow$ and $p_{23}\downarrow$. It is also credible that with $p_{31}\uparrow$ we may take $p_{33}\downarrow$ and $p_{32}\uparrow$, since faster maturation may result from $p_{31}\uparrow$ plus intensive timber improvement practices.

Our goal is to have $\vec{y}_0\left(\prod_{t=1}^{N} B_t\right)E \uparrow$ and bounded as $N \to \infty$ for a practical range of $\vec{y}_0$, when $t \geq T$. To get $\uparrow$ we need

$$\left(\vec{y}_0 \prod_{t=T}^{N+1} B_t - \vec{y}_0 \prod_{t=T}^{N} B_t\right)E \geq 0,$$

which occurs if $(B_{N+1} - I)E \geq 0$. For boundedness we may use $\left\|\prod_{t=1}^{N} B_t\right\|$ is bounded.

To illustrate our approach, consider the hypothetical case,

$$A_t = A + h_t J \quad \text{where} \quad A = \begin{pmatrix} 0.1 & 0.7 \\ 0.8 & 0.1 \end{pmatrix}$$

$$0 < h_t \to 0 \downarrow, \quad J = \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix}.$$

category as a result of growth during a transition period (see definition of cutting period below).

If a forest is not now providing a sustained yield, that is, it is not yielding a periodic harvest equal to its periodic growth maximized subject to environmental constraints and optimal silvicultural practices, then it is desirable to manage the forest by a long range plan whose purpose is to achieve sustained yield. There exist standard procedures applied by professional foresters. But to illustrate some of our "growth factor" theory we consider a different and very idealized approach which emphasizes a sustained growth stock.

Our goal is to stabilize the total growth stock. We assume the forest is an uneven-age type on which selective cutting (in contrast to clearcutting) is the principal method of harvest. We also assume that the present condition of the forest is one of inadequate growth stock and excess of mature and overmature trees, and capable of receiving management practice which if periodically applied will achieve a monotonic increase of growth stock to a silviculturally and ecologically desirable level.

$$A = \begin{pmatrix} 0.1 & 0.7 \\ 0.8 & 0.1 \end{pmatrix},$$

we check (*) as follows:

$$(0.1 - h_t)/\lambda \quad + \quad (0.7 + h_t)/\lambda \qquad 0.2 \geq \frac{0.1 - h_t}{0.9} + \frac{0.7 + h_t}{0.7} \quad ,$$

$$(0.8 - h_t)/\lambda \quad + \quad (0.1 + h_t)/\lambda \qquad 0.2 \geq \frac{0.8 - h_t}{0.9} + \frac{0.1 + h_t}{0.7} \quad .$$

If h_t is sufficiently small, both of these are ≥ 1. This will work with G also.

For boundedness, consider

$$\left\| \prod_{t=1}^{N} (A + h_t J) \right\| \leq \prod_{t=1}^{N} \|A + h_t J\| \leq \prod_{t=1}^{N} (\|A\| + 2\,|h_t|)$$

$$= \|A\|^N \prod_{t=1}^{N} \left(1 + 2\,\frac{h_t}{\|A\|}\right) \; .$$

So, if

$$\sum_{t=1}^{\infty} h_t < \infty \qquad \text{then} \quad \prod_{t=1}^{N}\left(1+2\,\frac{h_t}{\|A\|}\right)$$

converges. Now if we use A_tH instead of A_t, then

$$\left\|\prod_{t=1}^{N} A_tH\right\| \leq \|AH\|^N \prod_{t=1}^{N}\left(1+2h_t\,\frac{\|H\|}{\|AH\|}\right).$$

This will be bounded if $\|AH\| \leq 1$. But $H = \mathrm{diag}(\lambda^{-1}, |\mu|^{-1})$ will not work. Now we note that the conditions for boundedness and monotoneity together imply $AHE = E$. Since A^{-1} exists, we find

$$H \doteq \mathrm{diag}(1.09, 1.27).$$

For this H we verify $A_tHE \geqq E$ for all t. Hence both conditions are satisfied, so we have an example of the model.

V. REFERENCES

Bowerman, B., David, H. T. and Isaacson, D. 1975. Uniform strong ergodicities of non-homogeneous Markov chains. Preprint, Department of Statistics, Iowa State University.

Bowerman, B., David, H. T. and Isaacson, D. 1977. The convergence of Cesaro averages for certain non-stationary stochastic chains. Stochastic Proc. and their Appl. (To appear).

Breiman, L. 1969. Probability and stochastic processes. Boston, Houghton Mifflin.

Conn, P. S. 1969. Asymptotic properties of sequences of positive kernels. Ph. D. thesis, Iowa State University, Ames, Iowa.

Dobrushin, R. L. 1956. Central limit theorems for non-stationary Markov chains I. II. Theor. Probability Appl. 1: 65-80, 329-383.

Feller, W. 1949. Fluctuation theory of recurrent events. Trans. Amer. Math. Soc. 67: 98-119.

Gantmacher, F. R. 1959. Applications of the Theory of Matrices. Interscience Publishers, Inc., New York.

Hajnal, J. 1956. The ergodic properties of non-homogeneous finite Markov chains. Proc. Camb. Phil. Soc. 52: 66-67.

Isaacson, D. and Madsen, R. 1974. L_1 ergodic behavior of non-negative kernels. Israel Journal of Meth. 17: 350-363.

Isaacson, D. and Madsen, R. 1976. Markov Chains Theory and Applications. Wiley, New York.

Kabak, I. W. 1970. Wildlife management: An application of a finite Markov chain. Am. Stat. 24: 27-29.

Karlin, S. 1955. On the renewal equation. Pacific J. Math. 5: 229-257.

Lipstein, B. 1965. A mathematical model of consumer behavior. J. Marketing Res. 2: 259-265.

Madsen, R. and Conn, P. 1973. Ergodic behavior for non-negative kernels. Annals of Prob. 1, No. 6: 995-1014.

Madsen, R. and Isaacson, D. 1973. Strong ergodic behavior for non-stationary Markov processes. Annals of Prob. 1, No. 2: 329-335.

Mott, J. L. 1957. Conditions for the ergodicity of non-homogeneous finite Markov chains. Proc. of Royal Soc., Edinburgh 64: 369-380.

Ostrowski, A. M. 1960. On the eigenvector belonging to the maximal root of non-negative matrix. MRC Technical Summary Report. No. 145.

Paz, A. 1970. Ergodic theorems for infinite probabilistic tables. Annals of Math. Statist. 41: 539-550.

Paz, A. 1971. Introduction to Probabilistic Automata. Academic Press, New York.

Pitman, J. W. 1974. Uniform rates of convergence for Markov chain transition probabilities. Z. Wahr. Verw. Geb. 29; 193-227.

Pullman, N. and Styan, G. 1973. The convergence of Markov chains with non-stationary transition probabilities and constant causative matrix. Stochastic Proc. and their Appl. I: 279-285.

Timber management plan of the Dixie Natural Forest. 1976. U.S. Forest Service, Region 4. Cedar City, Utah.

VI. ACKNOWLEDGMENTS

I sincerely wish to thank my co-majors advisors Dr. D. Isaacson and Dr. B. Vinograde, for their extremely patient, able guidance in the research and writing of this thesis. Their constant encouragement and unselfish availability have been deeply appreciated. Both of these men sincerely believe in the worth of each individual, and it is a real pleasure to have been associated with them.

Thanks are also given to the other members of my committee, Professors R. Lambert, G. Meeden and R. Warren, for generously giving their time and support. Furthermore, I greatly appreciate the financial support given me by the Mathematics Department during the course of my graduate program at Iowa State University.

Also, I would like to thank Madonna Blanshan for her extremely able typing of this thesis.

I am grateful to my parents, Chung-Yi Huang and Hong-Kan Yang Huang, for their love during the years and for their encouragement to go to college. Also, my life has been made much happier by knowing my sister Ming-Chu and my brother Wen-Ching.

VII. APPENDIX.

POTENTIAL YIELD BY 20 YEAR PERIODS

PONDEROSA PINE WORKING GROUP - STANDARD COMPONENT

20 Year Cutting Period	Type of Harvest	Cut Acres	20 Year Yield Volume MMBF	Total
1	Regeneration	26,157	140	140
2	Regeneration	25,751	140	140
3	Regeneration	25,151	140	140

At this point intermediate harvest of intensively managed existing stands begins.

20 Year Cutting Period	Type of Harvest	Cut Acres	20 Year Yield Volume MMBF	Total
4	Regeneration	25,693	140	
	Intermediate	16,199	41	181
5	Regeneration	25,268	140	
	Intermediate	16,199	41	181

At this point begins intermediate cut on prior regeneration harvest areas.

20 Year Cutting Period	Type of Harvest	Cut Acres	20 Year Yield Volume MMBF	Total
6	Intermediate	26,157	66	
	Regeneration	15,349	140	206
7	Intermediate	25,751	65	
	Regeneration	15,774	140	205

At this point all stands are approximately rotation ages or younger but there is still an imbalance in acres by age class

20 Year Cutting Period	Type of Harvest	Cut Acres	20 Year Yield Volume MMBF	Total
8	Intermediate	25,151	63	
	Regeneration	22,706	241	304
9	Intermediate	25,693	65	
	Regeneration	22,706	241	306
10	Intermediate	25,268	64	
	Regeneration	22,706	241	305
11	Intermediate	15,349	39	
	Regeneration	22,706	241	280
12	Intermediate	15,774	40	
	Regeneration	22,706	241	281

At this point there is achieved a balance of acres in all age classes. Hereafter the 20 year potential yield will remain at 299 MMBM.

13	Intermediate	22,706	57	
	Regeneration	22,706	241	299
14	Intermediate	22,706	57	
	Regeneration	22,706	241	299